STRUCTURAL THEORIES OF MODELING TOKEN CAUSATION

Michael Joachim Stentenbach

LONDON SCHOOL OF ECONOMICS
AND POLITICAL SCIENCE

October 2006

thesis submitted to the
UNIVERSITY OF LONDON

for the completion of the degree of a
DOCTOR OF PHILOSOPHY

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

For my parents
who have always believed in me.

ABSTRACT

This thesis deals with the most prominent accounts of analyzing singular event causation by employing counterfactuals or counterfactual information. The classic counterfactual account of token event causation was proposed in 1973 by the philosopher David Lewis and ruled that an event *c* is a cause of event *e*, if and only if there is a chain of counterfactually dependent events between *c* and *e*. Apart from facing conceptual problems due to its metaphysical claim to analyze causation 'as such' and to reduce it to counterfactual dependency, this account also produced implausible results: first, it stipulated that token causation is a transitive relation, and second, it could not analyze situations in which an effect is over-determined by various causes, either symmetrically or by one cause pre-empting another one.

In 2000, almost three decades later, Judea Pearl, formerly an engineer, formulated a new and highly influential theory of modeling causal dependencies using counterfactual information that, as I argue, neither faces these conceptual problems nor produces these undesired results. This formal theory analyzes causal relationships between token events in a given situation in two steps: first, a causal model describing the relevant mechanisms at work in the situation is constructed, and second, causal relationships between the events featured in the situation are determined relatively to this model. Pearl's definition of causation according to a model is technically complicated, but its underlying rationale is that the decisive property of a cause is to sustain its effect via a certain causal process against possible contingencies, this notion of sustenance embodying an aspect of production and an aspect of counterfactual dependency.

This theory of Pearl's was received with great interest in the philosophical community, most importantly by Christopher Hitchcock and James Woodward, who tried to simplify this account while preserving the basic intuition that a cause is linked to its effect by a causal process, in essence a concatenation of the mechanisms at work in the situation, just defining a causal process in a formally simpler way.

I describe and employ this simplified account by Hitchcock and Woodward as a graphic introduction to Pearl's theory, because the same basic notions, like the one of a causal model, are defined in a formally more accessible way and the basic problems, like the generation of a causal model, become obvious. I mainly discuss Hitchcock's account, since this is the earlier one, since it is more elaborate, and mainly since it is conceptually in need of clarification. Woodward's account is in essence equivalent to Hitchcock's, given a slightly changed terminology.

The core of my thesis consists of a comparison of Pearl's theory with Hitchcock's account. I present four paradigmatic examples, three of which are judged differently by these two theories. In each of these three examples our causal intuition is in accord with the judgment delivered by Pearl's account but contradicts the verdict of Hitchcock's. I draw the conclusion that Hitchcock's project of simplifying Pearl's theory fails in the second step of causal analysis, i.e. in defining causation according to a given model.

Building on the lessons learned from this comparison, I offer a slight generalization of Pearl's definition of token causation according to a model, since Pearl's original account has the shortcoming that token causes cannot be exogenous in a model.

ACKNOWLEDGEMENTS

I am deeply indebted to many people without who this thesis would not have come together:

First to Professor Rosemarie Rheinwald who started my training in analytic philosophy and who woke my interest in the study of causation, after I had started out as a mathematician originally. Also to Dr Matthias Paul for suggesting to pursue my PhD research here and in the philosophy of science – I have never regretted this.

My main gratitude goes to Professor Nancy Cartwright for supervising this thesis. Without her, I would never have made the turn to understand philosophy as an applied meta-science, and it was her advice that made me investigate interventionist accounts of causation. All in all, this thesis is a result of the things I have learned from her. Moreover, I am deeply indebted to Professor Colin Howson for co-supervising this thesis and for having been such an inexhaustible source of encouragement and humor.

Special thanks go to Dr Jon Williamson for a year of fruitful collaboration and advice – this year marked the breakthrough in my understanding of causation. Further thanks go to Dr Damien Fennell for strong support during the process of wrapping up the thesis. Lastly, I would like to thank all the people in the *Department of Philosophy, Logic and Scientific Method* and the *Centre for the Philosophy of the Natural and the Social Sciences* for having made my time here such a pleasant experience. These years have formed not only my work but also a part of myself.

On the private side, I owe more than can be imagined to my parents. Not only have they always believed in me, it was them who have put me on the right track from the very beginning. So far, this has been the result.

Michael Stentenbach
London School of Economics, October 2006

TABLE OF CONTENTS

CHAPTER 2

PEARL'S ACCOUNT OF MODELING TOKEN CAUSATION 100

CHAPTER 3

PARADIGMATIC CASES JUDGED DIFFERENTLY 170

INTRODUCTION AND OVERVIEW

In 1973, in the article 'Causation', David Lewis proposed his classic counterfactual account of token causation that became one of the most debated views in the philosophical study of causation in the last decades.[1] In essence, this account ruled that for two occurring events c and e in a situation S, event c actually causes event e if and only if there is a chain of events $d_1,\ldots,d_n$, such that d_1 is causally dependent on c, d_2 is causally dependent on d_1, ... , and e is causally dependent on d_n. In this context, event d_{i+1} is causally dependent on event d_i if and only if the following counterfactual holds: 'If d_i had not occurred, then d_{i+1} would not have occurred.' Lewis gave an elaborate semantics of so-called possible worlds in order to account for the truth of these counterfactuals, and he even championed the highly controversial thesis that these possible worlds were real in a physical sense.[2] However, this claim of Lewis's to account for a metaphysical notion of causation and moreover to reduce this notion to counterfactual dependency, was more commonly perceived as a problem for this account rather than as an advantage.

The striking response that Lewis's account received was in my view due to its main feature, namely simply to describe causal relationships with the aid of counterfactuals. For, this mode of representation of token-level causal relationships offers benefits in various ways. On the one hand, formally, it is a remarkably simple way to describe causal dependencies. On the other hand, this mode of representation reflects at least part of our causal intuition, which often regards an event as causally responsible for another if they are counterfactually dependent. In my view, though,

[1] Cf. [dL 73].
[2] Cf. Lewis's book *Counterfactuals*, [dL 73b].

the main benefit of a counterfactual description of causal relationships lies in the fact that such a description offers information that could theoretically be exploited for forecasting and also of manipulating described events. It is this possibility of offering knowledge beyond a mere descriptive causal explanation, i.e. knowledge that can be applied in predictions and control, which makes a counter-factual description of causal relationships so interesting.[3]

Unfortunately, Lewis's account did not only have these rather abstract features, which it moreover shares with every counterfactual account on a foundational level and which make all of these accounts highly desirable. Lewis's account also produced implausible results. The following are in my view the basic three problems that have become apparent in the application of Lewis's account: First, Lewis's account stipulates that token causation is a transitive relation. Second, Lewis's account cannot analyze situations in which an effect is symmetrically over-determined by various causes. Third, Lewis's account can also not analyze situations in which an effect is asymmetrically over-determined by one cause pre-empting another one.

Failure of Lewis's classic account in these three points has been widely recognized in the philosophical community.[4] In his recent paper 'Two Concepts of Causation', Ned Hall attributed this failure partly to the fact that Lewis's account does not completely capture our pre-theoretic intuition about token causation.[5] According to Hall, token-level causal relationships have at least two aspects in our intuition, a production aspect, according to which a cause *c* has to bring about its effect *e*, and a dependence aspect, according to which *e* counterfactually depends on *c*. Lewis's account obviously only captures the latter aspect.

---

[3] Cf. also [jP 00] for further remarks on the applicability of counterfactual information for the purpose of prediction and control.

[4].Cf. for example the collection of articles in [CHP 04].

[5] Cf. [nH 04].

In 2000, Lewis reacted to these shortcomings of his classic account by gravely modifying it in his new article 'Causation as Influence'.[6] This new account still ruled that event c causes event e in a situation S if and only if there is a chain of events $d_1,\ldots,d_n$, linking c and e. However, the relation that had to hold between the links d_i and d_{i+1} of this chain was no longer counterfactual dependence, but the new relation of influence. In essence, d_i influences d_{i+1} according to Lewis if and only if there is a range of alterations (d_i', d_i'',...) of d_i and a range of alterations (d_{i+1}', d_{i+1}'',...) of d_{i+1} such that the following counterfactuals hold: 'If d_i' had occurred, then d_{i+1}' would have occurred.', 'If d_i'' had occurred, then d_{i+1}'' would have occurred.', ... This introduction of event-alterations into Lewis's new account was an improvement. However, Lewis's new account still inherited the shortcoming of its predecessor by stipulating transitivity of causation.[7] Moreover, it also introduced new problems, like counter-intuitively analyzing events that were intuitively spurious causes as proper causes.[8] All in all, this new account was not the solution hoped for. In my view, help came from another direction.

Also in 2000, in his book *Causality – Models, Reasoning, and Inference*,[9] Judea Pearl formulated a new counterfactual account of both the type-level and the token-level causation. Pearl started his study of causation from an empiricist tradition. Coming from an engineering background and not being a philosopher by training, Pearl mainly conducted research in probability theory and artificial intelligence and had the conviction that causal relationships were in essence useful abbreviations of probabilistic patterns. Pearl's main work in this field was his book *Probabilistic Reasoning in Intelligent Systems*.[10]

---

[6] Cf. [dL 00].
[7] Cf. for example [jC 00].
[8] Cf. for example [iK 01].
[9] Cf. [jP 00].
[10] Cf. [jP 88].

Since then, Pearl's view of causation has changed dramatically. In his book *Causality*, Pearl regards the probabilistic relationships as mere surface phenomena and takes the causal relationships as the fundamental building blocks both of physical reality and of our understanding of this reality.[11] *Causality* is now commonly regarded as a milestone in the study of causation, and without a doubt it has been more influential in the philosophical community than any other of Pearl's publications.

In this thesis, I will only deal with the token-level account of causal modeling that Pearl offers in *Causality*. I will argue that this account avoids Lewis's conceptual problem of attempting to analyze a meta-physical notion of causation. Furthermore, I will claim that there is evidence that this account of Pearl's also surmounts the three basic problems of Lewis's counterfactual account: symmetric and asymmetric over-determination and the stipulation of a transitivity of token causation.

Pearl's formal account analyzes causal relationships between token events in a given situation S in two steps: First, a causal model M describing the relevant mechanisms at work in situation S is constructed. Second, causal relationships between the events featured in situation S are determined relative to this model M. I will argue that this strategy of Pearl's to divide the problem of causal analysis into two sub-problems – first the generation of an appropriate model M formalizing situation S, and second the definition of causation in S relativized to such a model M – offers many advantages over the classic undivided approach: On the one hand this division trivially avoids striving to account for a metaphysical notion of causation that is independent of our description. On the other, it also allows accounting for the fact that there are various descriptions of the same situation S that are all equally justified, depending on which events, and also on which kinds of events, in S are of interest for us.

[11] Cf. the introduction of [jP 00].

Pearl's definition of token causation according to a model is technically complicated, but its underlying rationale is that the decisive property of a cause is to sustain its effect via a certain causal process against possible contingencies. I will argue that this notion of sustenance, embodying both an aspect of production and an aspect of counterfactual dependency in Hall's terminology, is responsible for the success of Pearl's account of modeling token causation.

Pearl's account of modeling token causation was received with great interest in the philosophical community, most importantly by Christopher Hitchcock in his article 'The Intransitivity of Causation Revealed in Equations and Graphs' and by James Woodward in his influential book *Making Things Happen*.[12] In my understanding, both Hitchcock and Woodward tried to simplify Pearl's account while preserving its basic rationale: In a situation S a cause c is linked to its effect e by a causal process, which is in essence a concatenation of the mechanisms at work in S. In my view, Hitchcock and Woodward wanted to achieve this simplification by discarding Pearl's notion of sustenance in the formalization of a causal process. Instead, they opted for identifying a causal process in a situation S with a route in a causal model M of this situation.

[12] Cf. [cH 01] and [jW 03].

This thesis follows the leading question whether this simplification of Pearl's account by Hitchcock and Woodward succeeds. In answering this question I proceed in the following way:

The first chapter is intended to give a thorough clarification and corrective reconstruction of Hitchcock's account of token causation which he gives in 'The Intransitivity of Causation Revealed in Equations and Graphs' and is also intended to briefly touch on the account that Woodward sketches in *Making Things Happen*. As the title of Hitchcock's article already suggests, his main intention, apart from offering a simplification of Pearl's account, is to motivate that token causation is not a transitive relation in general. In contrast to this, Woodward merely intends to give a sketch of certain characteristics that a successful account of token causation has to fulfill. For the details of this sketch Woodward explicitly refers to Hitchcock.
My reason for focusing on Hitchcock's account of token causation in this chapter, apart from the fact that Woodward explicitly refers to it, is first that it is more detailed and second that it is in strong need of a conceptual clarification. The argumentation that I will offer in this chapter represents my attempt to reconstruct Hitchcock's rather involved original in the most charitable and unambiguous way in order to find out how his account really functions.
The only point in which my corrective reconstruction deviates from Hitchcock's original is the way in which I relativize his definition of definitions of token causation to a model M of the underlying situation S. In my view, such a relativization allows us to analyze situations S from various viewpoints, depending on which events, and also which kinds of events, in S are of interest for us.

Starting this way with a thorough clarification of Hitchcock's account also allows the introduction of the basic notions of a counterfactual account of causal modeling in a simple and illustrative way, so that this chapter also serves as a preparation to the formally more advanced subsequent chapter which deals with Pearl's influential approach in modeling token causation.

As already mentioned, both Hitchcock and Woodward formulated their accounts under the influence of Pearl's, and in my view their main intention was to simplify Pearl's formally rather elaborate work by finding simpler formal expressions of the basic rationale in Pearl's work: The understanding that first a token cause *c* and its effect *e* in a situation *S* are linked by a causal process and second that such a process is in essence a concatenation of mechanisms at work in *S*.

I argue that Hitchcock formalizes a causal process between events *c* and *e* in a situation *S* by a route in a causal model *M* of this situation. Following Pearl, Hitchcock implicitly differentiates two kinds of token causation. The first is *actual causation* which applies to cases in which an effect *e* is caused by a single event *c*. The second is a weakened version of token causation which I dub *contributory causation* and which is intended to cover cases of symmetric over-determination in addition to the ones just mentioned. I argue that there is evidence that my reconstruction of Hitchcock's definitions of actual causation and of contributory causation can overcome the basic three problems of Lewis's counter-factual account – symmetric and asymmetric over-determination and the intransitivity of actual causation.

I round off this chapter by briefly discussing Woodward's account, which does not distinguish between actual and contributory causation. For Woodward, there is only one inclusive notion of singular event causation. However, I demonstrate that Woodward's definition of this inclusive notion of token causation is equivalent to my reconstruction of Hitchcock's definition of contributory causation.
Finally, I briefly touch on the problem of how Hitchcock's causal models *M* formalizing a given situation *S* can be constructed. I argue that this model generation procedure has to consist of two steps: First, the extraction of variables in situation *S*, and second the establishment of counterfactual dependencies between these variables and establishment of the structural equations.

In the second chapter, I focus on the account of modeling token causation that Pearl gives in his book *Causality*. Pearl gives accounts of many aspects of causal discourse in this book. He deals with causation at both the type and the token level, discusses the problem how to derive causal models from raw data and gives examples of plenty of applications in various fields of science. However, I limit my discussion of Pearl to in fact token causation, i.e. to the relation that holds between two singular events *c* and *e* in a situation *S*, when *c* in fact causes *e*, and also to the second problem of causal modeling, i.e. to the definition of token causation relativized to an appropriate causal model *M* formalizing a given situation *S*.
Pearl's basic rationale is that for an event *c* to qualify as a token cause for another event *e* in an arbitrary situation *S*, event *c* has to be able to sustain event *e* against certain contingencies. I will maintain that this rationale of Pearl's can be most easily understood with the aid of the leading idea that *c* sustains its effect *e* via a causal process that links cause *c* and effect *e*.

I argue that Pearl formalizes such a process by a causal beam, which is in essence a simplification of a given causal model M achieved by projecting M on the values of its exogenous variables. The leading idea for this projection is in broad terms for every variable Y in M to filter out the variables W in M that are not involved in sustaining the actual value of Y. Pearl accounts for two kinds of token causation, actual causation, applying to cases in which an effect e is caused by a single event c, and contributory causation, applying solely to cases of symmetric over-determination. I highlight a peculiar feature of his account, namely that exogenous variables can be neither actual nor contributory causes.

Again, I only marginally mention the first problem of causal modeling, i.e. the derivation of an appropriate causal model M from a given situation S. I give a brief synopsis of how Pearl generates his causal models and contrast this to the way in which Hitchcock by assuming total information about the situation S to be modeled tries to avoid the problem of model generation completely by in essence just stipulating a certain model. I briefly discuss Pearl's so-called algorithm of inductive causation and reach the result that this algorithm certainly offers criteria that a suitable model has to fulfill. However, it turns out that these criteria are insufficient to determine a model on which Pearl's definitions of actual and contributory causation can be applied.

In the third chapter, I construct four paradigmatic examples to which I apply my reconstructions of Hitchcock's definitions of actual and contributory causation, Pearl's corresponding definitions, and finally Woodward's definition of token causation. The application of these different accounts to my four examples serves various purposes: First, it makes it easy to compare the extensions of the different accounts, i.e. one can easily determine whether a cause according to one definition is also a cause according to another one. Second, the verdicts of these accounts can be compared not only to each other, but also with our causal intuitions in these examples, and this allows us to find out whether some definitions are more plausible than others. Third, possible formal difficulties in the application of these definitions become obvious.

The chapter has two leading questions: The first is whether Hitchcock, and in this way Woodward, too, succeed in giving a simplification of Pearl's account of singular event causation. I come to the conclusion that they do not. I show that my reconstruction of Hitchcock's definitions of actual and contributory causation analyzes three of my examples in a way that contradicts what I take to be our causal intuitions in these cases. My reconstruction of Woodward shows that his inclusive definition of token causation fares only slightly better in analyzing two of these examples incorrectly. In striking contrast to this, Pearl's account of actual and contributory causation, given that certain prerequisites are met, analyzes all four examples according to our intuition. From the discussion of these four examples I conclude that Hitchcock's and Woodward's project of simplifying Pearl's account fails because their underlying rationale, which identifies a causal process linking a cause and its effect with a route in a causal model, is too simplistic.

The second leading question is whether one can improve on Pearl's account. By applying Pearl's account of actual and contributory causation to these four examples, it becomes obvious that its inability to admit exogenous variables as causes is an unnecessary shortcoming. For this reason, I offer a natural extension of Pearl's account that facilitates a uniform treatment of exogenous and endogenous variables in a causal model and in this way also allow exogenous variables to be causes. Finally, I give evidence that this extension of mine can surmount the three major problems of Lewis's classic counterfactual account: the intransitivity of causation, preemption, and symmetric over-determination. I motivate this by demonstrating that my extended account can successfully analyze three examples featuring the respective problems.

CHAPTER 1

HITCHCOCK'S AND WOODWARD'S ACCOUNTS OF MODELING TOKEN CAUSATION

This chapter is intended to give a thorough clarification and corrective reconstruction of the accounts of token causation that Christopher Hitchcock offers in his article 'The Intransitivity of Causation Revealed in Equations and Graphs' and that James Woodward sketches in his influential book *Making Things Happen*.[13] In a certain way, this mode of presentation will do injustice to Woodward, because he does not really intend to give his own account of token causation. Only a comparatively small part of the book deals with modeling token causation. Instead, *Making Things Happen* gives an overview of an interventionist theory of causation and explanation. It touches on topics like the notion of an intervention and its justification, the notion of invariance and the way in which it offers a new understanding of the concept of a law in causal explanations, and the issue of how causal relationships can be exploited for purposes of manipulation and control.

Concerning the issue of modeling token causation, Woodward's intention is merely to give a sketch of certain characteristics that a successful account of token causation has to fulfill in his view. For the details of this sketch Woodward explicitly refers to Hitchcock's article.[14] My presentation and discussion of Woodward's sketch that I offer in this section is very short compared to the way in which I deal with Hitchcock. I see my justification for this mode of presentation in the following fact: The relevant definitions in Woodward's sketch of token causation are all taken over from Hitchcock. In this way, most of the remarks that I make about Hitchcock's account in this chapter equally apply to Woodward's account. My reason for focusing on Hitchcock's account of token causation in this chapter is mainly that Woodward explicitly refers to it, that it is more detailed, and finally that it is in strong need of a conceptual clarification.

---

[13] Cf. [cH 01] and [jW 03].

[14] Cf. for example [jW 03], p. 83.

In his article 'The Intransitivity of Causation Revealed in Equations and Graphs' Christopher Hitchcock proposes an analysis of singular event causation with the aid of a framework of structural equations. These structural equations are deterministic functions that encode 'complete' counterfactual information about the situation that is causally analyzed. Since Hitchcock's account is given in a very informal and suggestive way, this chapter is mainly devoted to clarifying how this account really functions. The argumentation that I offer in this chapter represents my attempt to reconstruct Hitchcock's account in the most charitable and unambiguous way. I merely refer to Hitchcock's original formulations and definitions without quoting them, and start out directly with my reconstructions. In this way, I hope to save the reader any unnecessary confusion that might arise when consulting Hitchcock's rather involved original. In the course of this chapter, I will work out the implicit assumptions underlying Hitchcock's methodology. Moreover, I try to give clear formal definitions of the concepts that Hitchcock, partially implicitly, employs. Thus the basic notions of a counterfactual account of causal modeling will be introduced in a simple and illustrative way, so that this chapter also serves as a preparation to the formally more advanced following chapter of this thesis that deals with Judea Pearl's influential approach in modeling token causation.

Both Hitchcock and Woodward formulated their accounts after having received the one of Pearl, and in my view their main intention was to simplify Pearl's formally rather elaborate work by finding simpler formal expressions of the basic intuitions in Pearl's work.[15] For this reason, I will compare the results of Hitchcock's and Woodward's accounts with the ones of Pearl's in the final chapter of the thesis in order to find out whether their simplifications succeed. In chapters two and three of this

[15] Hitchcock himself refers to Pearl's account as representing the current state of the art in the science of causal modeling; cf. pp. 273 and 274 of [cH 01].

thesis I will for the sake of simplicity treat my reconstructions of Hitchcock and Woodward in this chapter as if they were their own versions. My justification for this treatment is that I consider my reconstructions to be not only the straightforward formal expressions of their implicit rationale, but also to be as close as possible to the respective original. The only point in which my reconstruction deviates from their original accounts concerns the relativization of their definitions of causation to given models. I mentioned in the introduction of this thesis that token causal modeling can be understood as splitting the problem of analyzing token causation into two sub-problems: The first being the generation of a suitable causal model M to analyze a situation S, and the second being the definition of token causation according to such a given model. Hitchcock and Woodward implicitly define token causation in a situation S as a notion that is independent of our description of S by a model M. In my corrective reconstruction in this chapter, I will argue for a more pragmatic definition of token causation relativized to a model M. In my view, this will allow us to analyze situations S from various viewpoints, depending on which events, and also which kinds of events, in S are of interest for us.

The first section will deal with the basic building block of Hitchcock's account of token causation, his definition of a causal model. In my understanding, Hitchcock's causal models M are in essence devised to analyze a given situation S that contains certain token events c, d, e, etc., in whose causal relationships we are interested. For this purpose, a causal model M is comprised of a set of variables V that formalize alterations of these events and of a set of structural equations E that is devised to determine the values of these variables. The underlying rationale is that these equations E formalize deterministic mechanisms, or local laws, that are at work in the situation S and that govern the corresponding events.

In my opinion, the procedure by which Hitchcock models a situation in order to analyze its causal structure has to consist of two steps: First, the extraction of variables V in situation S, and second the establishment of counterfactual dependencies between these variables V and establishment of the structural equations E. From this reconstruction of Hitchcock's model generation procedure, I will draw the conclusion that, although Hitchcock does not explicitly mention this, his structural equations are in essence another notation for exactly the counterfactual information that we put into the analysis of a situation S.

The second section will deal with Hitchcock's definition of actual causation. In my view, the basic idea underlying this definition is that c is a cause of e in a situation S if and only if there is a causal process linking these events c and e. I understand such a causal process as a concatenation of the mechanisms that are at work in the given situation S and that are linking events c and e. I will reconstruct Hitchcock's notion of actual causation in my definition (H AC). Moreover, I will argue that Hitchcock formalizes a causal process between c and e by an active route in a causal model M between the variables X and Z whose actual values x and z formalize these events c and e. This concept of an active route will be reconstructed by my definition (H AR). Both (H AR) and (H AC) deviate from Hitchcock's original formulation in an important respect: Hitchcock wants to account for a notion of causation 'as such' that is independent of our description, and he does not relativize his original definition of actual causation to a given model, whereas I maintain that such an understanding is misguided. However, I will postpone my argumentation for this to the next but one section of this chapter.

In the third section, I will illustrate the functioning of my reconstruction of Hitchcock's definitions (H AR) and (H AC). Concretely, I will discuss how they are applied to analyze two of Hitchcock's favorite examples, the first of which illustrates that actual causation is not a transitive relation in general and the second features a case of asymmetric over-determination. I will argue that for both examples the analysis by (H AR) and (H AC) is in accord with what I take to be our causal intuitions. Furthermore, I will illustrate that Lewis's classic counterfactual account fails in analyzing these examples. In this way, the conjecture that Hitchcock's account, reconstructed by (H AR) and (H AC), can surmount two of my featured three problems of the classic counterfactual account – pre-emption and the intransitivity of actual causation – gets some evidence.

In the fourth section, I will argue that the differentiation of the problem of analyzing token causation into two sub-problems – the generation of an appropriate causal model M for a situation S and the definition of causation according to a given model M – that is induced by a causal modeling account has a great advantage over classic metaphysical theories of causation: If we qualify the formal analysis of causation in situation S to a given causal model M, this analysis can account for the fact that there is a multitude of possible descriptions of this situation that may all be equally justified, depending on which kinds of events are of interest in S. Moreover, I will briefly discuss the remarkable attempt of Hitchcock's original account to conceptually reduce actual causation to a special kind of counterfactual, a so-called 'ENF-counterfactual', by which he wants to express the activity of a route. I will argue that this reduction falls short, since the concept of a surgical intervention has to enter in definition (H AR) and Hitchcock does not give any clues how an intervention could be expressed by counterfactuals.

The fifth section will deal with Hitchcock's account of contributory causation for cases of symmetric over-determination. Admittedly, the term 'contributory causation' is my terminology, since Hitchcock does not give this account a special name. He just states that his previous definition of actual causation, reconstructed by my (H AC), is not capable of analyzing cases appropriately in which an effect has two or more causes that work in parallel. For cases of this form, Hitchcock introduces a definition of a weakly active route, which I will reconstruct as (H WAR), and his definition of token causation for these cases, which I dub *contributory causation* and which I reconstruct as (H CC). In my reconstruction of Hitchcock, an event c is then a contributory cause of another event e in situation S according to a certain model M if and only if there is a weakly active route between the variables formalizing these events in this causal model. Again, Hitchcock himself does not relativize his original account of token causation for cases of symmetric over-determination to a given model M, but wants to account for causation independently of our description of a situation S. However, referring to my argumentation before, I will continue to maintain that such an understanding is misguided. As a point of application and comparison, I will show how my reconstructions of Hitchcock's definitions (H WAR) and (H CC) can be successfully applied in an analysis of Lewis's classic example of symmetric over-determination, the Firing Squad. Lewis famously claimed to have no clear intuition in this example, and as a conclusion I will show that Lewis's account consequently fails to analyze it.

In the sixth section I will give a brief synopsis of the basic facts of Woodward's account of token causation, where the term 'token causation' is not Woodward's original terminology, but is chosen by me to avoid confusion. In strong contrast to Hitchcock, Woodward does not distinguish between actual and contributory causation. For Woodward, there is only one inclusive notion of singular event causation that also applies to cases of symmetric over-determination. However, Woodward's definition of token causation (W TC) is equivalent to Hitchcock's definition of contributory causation (H CC) employing the notion of a weakly active route defined by (H WAR). The difference from Hitchcock is that Woodward does not limit the extension of his definition (W TC) to cases of symmetric over-determination, as Hitchcock does, but that Woodward takes his definition to cover all cases of singular event causation. Since I consider my remarks about the relevant concepts in Hitchcock's account of token causation to translate into results about the analogous concepts in Woodward's account, I mainly aim to show that the concepts Woodward employs are equivalent to Hitchcock's.

I will confine myself to a short demonstration that Woodward's definition of token causation (W TC) is equivalent to my reconstruction (H CC) of Hitchcock's definition of contributory causation. For this aim I will also mention a preliminary version of this definition that Woodward gives for heuristic reasons, and I will argue that this preliminary version is equivalent to my reconstruction (H AC) of Hitchcock's definition of actual causation. Admittedly, Woodward does not explicitly deal with the generation of causal models for analyzing token causation, and moreover he avoids precise formal definitions of most of these concepts. This is due to the fact that the modeling of token causation is only of minor importance for him. Nevertheless, I maintain that his understanding of the respective notions can be extracted from his discussion more or less straightforwardly.

HITCHCOCK'S RATIONALE OF FORMALIZING SITUATIONS BY DETERMINISTIC CAUSAL MODELS CONTAINING STRUCTURAL EQUATIONS

This section deals with the basic building block of Hitchcock's account of token causation, his definition of a causal model. Unfortunately, Hitchcock's description of his modeling framework is rather unclear. For this reason, this section and great parts of this chapter are devoted to a systematic clarification of Hitchcock's account. In particular, my formulation in this section just represents my attempt reconstruct Hitchcock's modeling framework in the most charitable way.

In my understanding, Hitchcock's causal models M are in essence devised to analyze a given situation S that contains certain token events c, d, e, etc., in whose causal relationships we are interested. For this purpose, these causal models M are comprised of variables V that are either exogenous or endogenous. Exogenous variables formalize admissible alterations of background events, or background circumstances, in the situation S, whose development is unknown to us. Analogously, endogenous variables formalize admissible alterations of the events in situation S, about whose development we are not ignorant. A set of structural equations E in the causal model M is devised to determine the values of exactly these endogenous variables from the values of the exogenous ones. The underlying rationale is that these equations E formalize deterministic mechanisms, or local laws, that are at work in the situation S and that govern the corresponding events.

I will reconstruct Hitchcock's modeling framework with the aid of three definitions and try to clarify the exact procedure according to which

causal models and structural equations have to be set up following Hitchcock. The first of these is my definition of a causal model (H CM), which is very close to Hitchcock's original and only differs in notation. The second is my definition of counterfactual dependence (H CD), and it is a more significant deviation from Hitchcock's original. My final definition of exogeneity, endogenity and structural equations (H EESE) does not have an explicit correlate in Hitchcock's text and has to be extracted from his description.

In my opinion, the procedure by which Hitchcock models a situation in order to analyze its causal structure has to consist of two steps: First, the extraction of variables V in situation S, and second the establishment of counterfactual dependencies between these variables V and establishment of the structural equations E. I will argue that the first step in essence consists of an individuation of the relevant events and their alterations in situation S and that this individuation depends on our interests in this situation. Furthermore, I will claim that in the second step only our pre-theoretic judgment about the described situation S enters, which contains the complete counterfactual information about the previously identified events and their possible alterations.

From this reconstruction of Hitchcock's model generation procedure, I will draw the conclusion that, although Hitchcock does not explicitly mention this, his structural equations are in essence another notation for exactly the counterfactual information that we put into the analysis of a situation S.

For Hitchcock, the causal analysis of a given example situation begins with the construction of an appropriate causal model formalizing the relevant information known about the situation. The following is my reconstruction of Hitchcock's definition of such a causal model. Admittedly, my formulation slightly deviates from Hitchcock's original. However, apart from a slightly changed notation, the only difference is that the following (H CM) is slightly more elaborate than Hitchcock's original formulation.[16]

Definition (Causal Model, Hitchcock) **(H CM)**

A *causal model* is an ordered pair $\langle V,E\rangle$, where V is a set of variables and E is a system of structural equations relating the values of the variables in V.

Variables in V formalize possible event-alterations in the situation S and are either *exogenous* or *endogenous* in the model. Correspondingly, the set of variables V can be divided into two disjoint subsets $\boldsymbol{U}$ of endogenous variables and W of endogenous variables, such that $V = \boldsymbol{U} \cup W$.

Analogously, the set of *structural equations* E falls into two subsets, E_U and E_W. The equations E_U for exogenous variables $\boldsymbol{U}$ have the form $U = id_U$ and state the value u of the exogenous variable U in the situation S. The equations E_W for endogenous variables W have the form

$$W = f_W(X_1, \ldots, X_n)$$

and express the value of endogenous variable W as a function of the variables X_i that can be either exogenous or endogenous.

[16] Cf. [cH 01], p. 280 for Hitchcock's original formulation.

Admittedly, in this characterization of a causal model, the concepts of exogenous and endogenous variables and also of structural equations are just mentioned and not defined. The proper definition of these concepts can be found in the next but one definition (H EESE). But before that, let me briefly try to motivate the underlying idea of this definition and also introduce some other helpful concepts.

In my understanding of Hitchcock, in a causal model M, the exogenous variables $\boldsymbol{U}$ formalize the admissible alterations of the background events in the situation S, whose development is unknown to us. Analogously, I take it that the endogenous variables W in M formalize admissible alterations of the events in situation S, about whose development we are not ignorant. Furthermore, in my understanding of Hitchcock, every admissible instantiation of values $\boldsymbol{u}$ for the exogenous variables $\boldsymbol{U}$ uniquely determines the values w of the endogenous variables W in M with the aid of the structural equations E. With this, Hitchcock's causal models are completely deterministic.[17]

In this context, I call an instantiation of the exogenous variables $\boldsymbol{U} \subseteq V$ of the causal model $M = \langle V, E \rangle$ the *state* of this model, and I call an instantiation of all variables V that is consistent with the structural equations E a *solution* of this model. In this way, a solution is brought about by letting the exogenous variables $\boldsymbol{U}$ take on certain values $\boldsymbol{u}$ which then determine the values w for the endogenous variables $W \subseteq V$, so that w can simply be computed out of $\boldsymbol{u}$ with the aid of the structural equations.[18]

According to Hitchcock, the structural equations E_W for endogenous variables W encode counterfactuals.[19] To be precise, in my reconstruction

---

[17] Hitchcock is not completely explicit about this property of his causal models to be uniquely determined by their structural equations and the values of their exogenous variables in his original description of his modelling framework. Cf. [cH 01], pp. 279-281. However, I consider this unique determinacy to be the straightforward reconstruction of Hitchcock's modelling framework and moreover to be implicitly intended by Hitchcock.

[18] Both of these concepts are introduced by me and not Hitchcock's original terminology. I just borrow these terms from Pearl for a useful abbreviation.

[19] Cf. [cH 01], p. 281.

of Hitchcock, the structural equations E in the model M are constructed according to the following definition of counterfactual dependence:[20]

Definition (Counterfactual Dependence, Hitchcock) **(H CD)**

Let $V = \{Y,X_1,\ldots,X_n,W\}$ be the set of variables selected to be relevant for the analysis of causal dependencies in a given situation S.

Then W is *counterfactually dependent* on Y if and only if there are values $y \neq y'$ of Y and $w \neq w'$ of W, and values x_i for the other variables X_i, i={1,...,n}, such that

$Y = y \;\&\; X_1 = x_1 \;\&\; \ldots \;\&\; X_n = x_n \;\boxed{\rightarrow}\; W = w$ and (CI)

$Y = y' \;\&\; X_1 = x_1 \;\&\; \ldots \;\&\; X_n = x_n \;\boxed{\rightarrow}\; W = w'$.

Here, the symbol '$\boxed{\rightarrow}$' denotes counterfactual implication, so that clause (CI) is short for 'If it were the case that $Y=y$ & $X_1=x_1$ &... & $X_n=x_n$, then it would be the case that $W=w$.'.

Continuing in my reconstruction of Hitchcock, structural equations are constructed according to the following definition, which also formally clarifies the notions of exogeneity and endogeneity of variables:[21]

Definition (Exogeneity, Endogeneity, Structural Equations, Hitchcock) **(H EESE)**

Let V be the set of variables selected to be relevant for the analysis of causal dependencies in a given situation S and let Y be an arbitrary variable in V.

Then the *structural equation* E_Y for variable Y is a function that gives the value y of Y as a result of all and only the variables $X_i \in V$ in its argument on which Y depends counterfactually.

---

[20] Cf. [cH 01], p. 280 for Hitchcock's original definition.

[21] This definition has no literal original in Hitchcock's text, but it can be condensed out of his original description. Cf. [cH01], pp. 279-284 for this.

If there is no variable $X_i \in \boldsymbol{V}$ so that Y counterfactually depends on X_i, then Y is called *exogenous*, and E_Y has the form $Y = y$, just stating the value y of Y.
If there are variables $\{X_1,\ldots,X_n\} \subseteq \boldsymbol{V}$ so that Y counterfactually depends on these X_i, then Y is called *endogenous*, and E_Y encodes a set of counterfactuals of the form

$$X_1=x_1 \;\&\; \ldots \;\&\; X_n=x_n \;\Box\!\!\rightarrow\; Y = f_Y(x_1,\ldots,x_n)$$

for all possible combinations of values $(x_1,\ldots,x_n)$ for the variables $X_1,\ldots,X_n$.

In the following, I will try to illustrate the precise way in which these above definitions, in particular (H EESE), function in modeling a given situation S. For this purpose, I will employ one of Hitchcock's favorite examples, his so-called 'Boulder' example, with which he argued for the intransitivity of causation. But before I do so, allow me a few further remarks.

Observe that the above definition (H EESE) induces a causal order or hierarchy in the model $\boldsymbol{M}$. If a variable Y counterfactually depends on other variables $\{X_1,\ldots,X_n\} \subseteq \boldsymbol{V}$, then Y is on a higher level than these X_i in the model $\boldsymbol{M}$ in the sense that Y is more remote from the exogenous variables in the model.[22] Following Hitchcock, I call these X_i *parents* of Y in this context and conversely Y a *child* of these X_i. In order to expand this terminology, let me call Z an *ancestor* of Y if and only if there are variables $X_1,\ldots,X_n \in \boldsymbol{V}$ such that Z is a parent of X_1, X_i is a parent of X_{i-1}, and Y is a parent of X_n; conversely, I call Y a descendant of Z if the same condition is fulfilled. Ideally, the structural equations E in the model $\boldsymbol{M}$ are ordered by recursion, so that every variable Y can be traced back to exogenous variables and does not have itself among its ancestors, in which case Y

[22] Note that the X_i can be either exogenous or endogenous, whereas Y is endogenous by definition.

would constitute a loop in the model. Hitchcock does not explicitly require such a recursive ordering of the equations for his causal models, but in the example cases he discusses his models all have this property.[23] Be this as it may, let me assume in the following that the structural equations in Hitchcock's causal models are recursively ordered. With this, it is particular ensured that Hitchcock's causal models are deterministic in the sense that every possible state of the model $\boldsymbol{M}$ uniquely determines a solution of this model. Note finally that these requirements of recursive ordering and determinism pose considerable constraints on the set of counterfactuals, out of which the structural equations are constructed according to definitions (H CD) and (H EESE).

Let me also remark that Hitchcock's causal models as defined by (H CM) can be graphically represented. The variables in $\boldsymbol{V}$ then form the nodes of a graph. An arrow is drawn from node X to node W if and only if W counterfactually depends on X according to definition (H CD). A causal graph does though not specify the values of the variables. This information is only found in the system of structural equations $\boldsymbol{E}$.

In order to do justice to Hitchcock, let me finally emphasize that the above definitions (H CM), (H CD) and (H EESE) are just my reconstructions of his account. My definition of a causal model (H CM) is very close to Hitchcock's original and only differs in notation. However, the other two definitions are more significant deviations from Hitchcock's original. (H EESE) does not have an explicit correlate in Hitchcock's text and has to be extracted from his description.[24] And definition (H CD) reads rather different in Hitchcock's original. To illustrate this, let me briefly state this definition of counterfactual dependence in Hitchcock's original words:[25]

---

[23] Cf. [cH 01], p. 281 for Hitchcock's own comments about this.
[24] Cf. again [cH01], pp. 279-284 for this.
[25] Cf. [cH 01], p. 280.

Definition (Counterfactual Dependence, Hitchcock's Misleading Original) **(HMO CD)**

W depends counterfactually on *Z* if and only if there are values *z* and *z'* of *Z* and other variables $X_1, \ldots, X_n$ with respective values $x_1, \ldots, x_n$, such that $f_W(z, x_1, \ldots, x_n) \neq f_W(z', x_1, \ldots, x_n)$.

I dismissed this formulation because of its striking circularity. Hitchcock tries to set up his structural equations with the aid of this definition of counterfactual dependence, but then presupposes exactly these structural equations f_W in (HMO CD).

Be this as it may. Let me now introduce Hitchcock's 'Boulder' example, in order to illustrate the functioning of my reconstructions (H CM), (H CD) and (H EESE):[26]

Boulder: 'A boulder is dislodged, and begins rolling ominously towards a hiker. Before it reaches him, the hiker sees the boulder and ducks. The boulder sails harmlessly over his head with nary a centimeter to spare. The hiker survives his ordeal.'

According to our pre-theoretic causal intuition, Hitchcock supposes, and I agree, that the falling of the boulder causes the hiker to duck, and the ducking of the hiker causes his survival, but the falling of the boulder does not cause the survival of the hiker.

Illustrating definition (H CM), and also following Hitchcock's original description, this example can be modeled as follows:[27] There are three bi-valued variables, *B*, *D* and *S*. Variable *B* is exogenous and has the following meta-assignment of events to its values:

[26] Cf. [cH 01], pp. 276 and 277 for Hitchcock's exposition of this example.
[27] Cf. [cH 01], pp. 295-298 for Hitchcock's discussion of this example.

$$B = \begin{cases} 0, & \text{if the boulder does not roll} \\ 1, & \text{if the boulder does roll} \end{cases}$$

D and S are endogenous and have the following meta-assignment of events to their values:

$$D = \begin{cases} 0, & \text{if the hiker does not duck} \\ 1, & \text{if the hiker does duck} \end{cases}$$

$$S = \begin{cases} 0, & \text{if the hiker does not survive} \\ 1, & \text{if the hiker does survive} \end{cases}$$

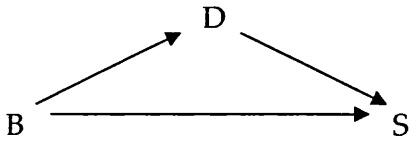
Illustrating definitions (H CD) and (H EESE), the structural equations for the endogenous variables D and S can be given in exhaustive form, stating the value of the image variable resulting from any possible combination of values for the argument variables:[28]

$$D = \begin{cases} 0, & \text{if B} = 0 \\ 1, & \text{if B} = 1 \end{cases} \quad \text{and} \quad S = \begin{cases} 0, & \text{if B} = 1 \text{ and D} = 0 \\ 1, & \text{if B} = 1 \text{ and D} = 1 \\ 1, & \text{if B} = 0 \text{ and D} = 0 \\ 1, & \text{if B} = 0 \text{ and D} = 1 \end{cases}$$

This notation suggestively illustrates the nature of Hitchcock's structural equations of endogenous variables – they are functions that state the value of the endogenous variable on the left hand side for every combination of values for the variables on the right hand side.

In the actual situation, B takes the value 1. Entering this value in the structural equations then yields the actual solution of this causal model: $B = 1$, $D = 1$, and $S = 1$.

The graphical representation of 'Boulder' finally is:



[28] This notation is my invention. For Hitchcock's original equations cf. [cH 01], p. 295.

With the aid of this example, I will in the following try to clarify the exact procedure according to which causal models and structural equations have to be set up following Hitchcock. In particular, I will try to reveal the dependence of the structural equations on underlying conditional information. In my opinion, the procedure by which Hitchcock models a situation in order to analyze its causal structure has to consist of the following two steps. The first of these illustrates the application of definition (H CM), and the second puts definitions (H CD) and (H EESE) in concrete terms:

M1) Extraction of variables V in situation S.

M2) Establishment of counterfactual dependencies between these variables V and establishment of the structural equations E.

In step M1), the relevant events and their possible alterations in situation S have to be identified with variables and their value-assignments. This extraction of variables is not a process of translation that is guided by grammatical rules.

In Hitchcock's example 'Boulder' the events of dislodging the boulder, or not, and of the boulder hitting the hiker, or not, could as well be regarded as demanding representation by a variable. For, there is no grammatical distinction between the dislodging of the boulder and the rolling of the boulder – both events are given in the description by predicates applying to the same subject. In particular, no conditionals or counterfactuals enter in this step of extracting variables. For, observe that the cover story of the example 'Boulder' does not even contain a conditional, let alone a counterfactual.

Instead, the extraction of variables and their values seems to be determined mainly by our pre-theoretic understanding and judgment of the situation regarding which events are relevant and which are not. In my view, such a pre-theoretic notion of relevance could for example be

understood as our naïve judgment whether the mentioning of a certain event is essential in a correct description of the situation under consideration. Such a description could in particular be a natural language description in the style of the cover story given with the example.
Furthermore, the individuation of event-alterations in situation S also heavily depends on our interests in the situation. For example, we could be interested in the velocity and the trajectory of the boulder. In that case we would take different possible velocities and trajectories as alterations of the actual event – the actual speed and path of the boulder. And we could also be interested in how exactly the boulder hits the hiker, in case it does. Then we would take different possible sites of impact, for example chest, head, legs, partial missing of the hiker, complete missing of the hiker, etc. as alterations of the actual event – the boulder missing the hiker. All these different evaluations of the situation S yield different causal models M. And in my view, these models M of situation S are all equally justified, as equally justified as are our prior interests in the situation S that lead to their generation. I will discuss this topic of a dependency of a causal model on our interests in the situation to be modeled more elaborately in the fourth section of this chapter.

Step M2) in the modeling procedure is to formally represent the conditional dependencies in the example situation with the aid of the variables V that are identified in step M1). It is instructive to give first a simplified non-technical description of how counterfactual dependence is established before discussing the general method that determines the structural equations. Start now with the more graphic description of establishing counterfactual dependence:
Consider an arbitrary variable Y and then test it for counterfactual dependence on every other variable X_i, $i = 1,\ldots,n$ in the causal model. I.e. begin with X_1 and find out whether there is a distribution of values for $X_2, \ldots, X_n$, such that a change in the value of X_1 brings about a change in

the value of Y, then repeat this procedure for the other X_i. If Y is counterfactually independent of all $X_i \in V$ in this sense, i.e. for all X_i there is no distribution of values to the other $X_1,\ldots,X_{i-1},X_{i+1},\ldots,X_n$, such that a change in the value of X_i would bring about a change in the value of Y, then Y is exogenous. Otherwise, Y is endogenous.

In this step of determining counterfactual dependencies between the variables in the model only our pre-theoretic judgment about the described situation S enters. This judgment though contains the complete conditional and counterfactual information about possible alterations of the previously identified events considered to be relevant in this situation. More concretely, this counterfactual information comprises all possible combinations of event alterations in the situation S and can most easily be represented by a set of counterfactuals. I call this set the *basic set of super-exhaustive counterfactuals.* For every variable $Y \in V$ this set contains a complete list of counterfactuals of the following form: For every possible combination of values of the variables in $V \setminus Y$ there is a counterfactual that has this combination as antecedent and that has the value y of Y corresponding to this combination in its consequent. Here, whether a combination of values of the variables in $V \setminus Y$ is possible, and whether a value y of Y corresponds to such a combination depends entirely on our pre-theoretic judgment of the situation.

For a deeper understanding of step M2) in the modeling procedure, it is instructive to describe this set of super-exhaustive counterfactuals for a variable Y more formally. For this purpose, assume without loss of generality that $V = \{Z,X_1,X_2,Y\}$. Suppose that Z can take k values $z_1,\ldots, z_k$, X_1 can take r values $x_{1,1},\ldots, y_{1,r}$, X_2 can take m values $x_{2,1},\ldots, x_{2,m}$ and that y can take p values $y_1,\ldots, y_p$ with $p < k \cdot r$. This means that there are $k \cdot r \cdot m$ possible value combinations for the antecedent variables, such that the set of super-exhaustive counterfactuals for variable Y contains the following $k \cdot r \cdot m$ counterfactuals:

For all $i \in \{1,\ldots,k\}$ and all $j \in \{1,\ldots,r\}$, there are m counterfactuals

$$Z = z_i \,\&\, X_1 = x_{1,j} \,\&\, X_2 = x_{2,1} \;\boxminus\!\!\rightarrow\; Y = y_{q1} \quad \text{for } q_1 \in \{1,\ldots,p\},$$

...

$$Z = z_i \,\&\, X_1 = x_{1,j} \,\&\, X_2 = x_{2,m} \;\boxminus\!\!\rightarrow\; Y = y_{qm} \quad \text{for } q_m \in \{1,\ldots,p\}.$$

It is this set that completely determines the structural equation for variable Y according to definitions (H CD) and (H EESE). To be more precise, the structural equation for variable Y can be condensed out of this set, if we remove all the variables in the antecedent on which Y is not counterfactually dependent. In the following, I describe how this elimination of redundant variables works.

Suppose for example that Y does not counterfactually depend on X_2 according to definition (H CD). Then for all m counterfactuals in the list above, the variable Z has the same value, i.e. it is $y_{q1} = \ldots = y_{qm}$. Then, for all $i \in \{1,\ldots,k\}$ and all $j \in \{1,\ldots,r\}$, these m counterfactuals above can be reduced to one counterfactual in which variable X_2 is omitted:

$$Z = z_i \,\&\, X_1 = x_{1,j} \;\boxminus\!\!\rightarrow\; Y = y_q \quad \text{for } q \in \{1,\ldots,p\}.$$

By this procedure, we have not only reduced the number of counterfactuals with Y in the consequent to k·r, but in particular we achieved that these resulting new counterfactuals now only contain the parents of Y in the antecedent. I call these k·r counterfactuals with this property the *exhaustive counterfactuals for the variable* Y and denote the set that consists of them the *set of exhaustive counterfactuals for the variable* Y.[29]

As a final point of my description of step M2) of the modeling procedure, let me highlight the connection between this set of exhaustive counterfactuals for Y and the structural equation for Y. The structural equation f_Y for variable Y encodes these k·r counterfactuals in the

[29] This procedure also works, if Y is exogenous. In this case, Y does not counterfactually depend on any of the Z, X_1, and X_2, so that the set of super-exhaustive counterfactuals reduces to the empty set $\varnothing$ as set of exhaustive counterfactuals for Y.

following way: For all $i \in \{1,\ldots,k\}$ and all $j \in \{1,\ldots,r\}$, the exhaustive counterfactual $Z = z_i$ & $X_1 = x_{1,j}$ $\square\!\!\rightarrow$ $Y = y_q$ translates into one line of the definiens of the structural equation f_Y in exhaustive form.

Let me illustrate this with the aid of the example 'Boulder' and the equation of variable S: The exhaustive counterfactual $B=1$ & $D=0$ $\square\!\!\rightarrow$ $S=0$ for example translates into the first line of the definiens of the structural equation f_S in exhaustive form, given by $S = \begin{cases} 0, & \text{if B = 1 and D = 0} \\ 1, & \text{if B = 1 and D = 1} \\ 1, & \text{if B = 0 and D = 0} \\ 1, & \text{if B = 0 and D = 1} \end{cases}$.

This exhaustive notation can then be condensed to a more practical form for this structural equation, so that we yield $S = \begin{cases} 1, & \text{if B = 0 or D = 1} \\ 0, & \text{otherwise} \end{cases}$.

The moral that I would like to draw from this attempt of mine to clarify Hitchcock's model generating procedure and to reconstruct it by steps M1) and M2) above is the following: Although Hitchcock does not explicitly mention this, his structural equations are in essence another notation for exactly the counterfactual information that we put into the analysis of a situation S in step M2). The complete information about situation S is already contained in the set of super-exhaustive counterfactuals that is stipulated at the beginning of the modeling procedure. In this way, structural equations do not generate new information, they just rewrite the information about the situation at hand that we either previously had available or that we previously stipulated. Trivially, this information that we stipulate about situation S in step M2) is about the events and their alterations that we individuate in step M1). And according to my observation in step M1), this individuation depends on our interests in this situation.

HITCHCOCK'S DEFINITION OF ACTUAL CAUSATION EMPLOYING ACTIVE ROUTES AS THE FORMALIZATION OF CAUSAL PROCESSES

This section will deal with Hitchcock's formal definition of actual causation. I will mainly discuss my reconstruction of Hitchcock's account. My formulation here and in great parts of this chapter just represents my attempt to clarify Hitchcock's account in the most charitable and unambiguous way.

In my view, the basic idea underlying this definition is that *c* is a cause of *e* in a situation *S* if and only if there is a causal process linking these events *c* and *e*. Here, I understand a causal process as a concatenation of the mechanisms that are at work in the given situation *S* and that are linking events *c* and *e*. I will argue that Hitchcock formalizes a causal process between *c* and *e* by an active route in a causal model *M* between the variables *X* and *Z* whose actual values *x* and *z* formalize these events *c* and *e*. In essence, an active route formally filters out a chain of mechanisms linking *c* and *e*, so that an alteration of event *c* would propagate along this chain to change the event *e* when the mechanisms in situation *S* that do not belong to this causal chain are screened off.

Three definitions form the core of my reconstruction of Hitchcock's account of actual causation. The first is Hitchcock's original definition of counterfactual dependence of one value of a variable on another value of another variable in a certain solution of a causal model, which I label (H CDSCM). The second definition is my reconstruction of Hitchcock's notion of an active route in a certain solution in a causal model, which I formulate in (H AR). Its leading idea is that a route $(X,Y_1,\ldots,Y_n,Z)$ in

model M between variables X and Z, whose values x and z formalize the events c and e of situation S, is active in the actual solution of M if and only if the following holds: A change of the value of X propagates exactly through the variables $Y_1,\ldots,Y_n$ in the route and changes the value of Y. This route then corresponds to a concatenation of mechanisms $E_X \circ E_{Y1} \circ \ldots \circ E_{Yn} \circ E_Z$ in situation S that links events c and e. The third definition finally is my reconstruction of Hitchcock's definition of actual causation, which I denote (H AC). It rules that event c is an actual cause of e in situation S according to the model M if and only if in the actual solution of M there is an active route from X to Z.

My reconstructions (H AR) and (H AC) of Hitchcock's account deviate from Hitchcock's original formulation in two respects: First, Hitchcock's original formulation neglects that the fact whether a route is active in a model M or not depends on the actual solution of this model M. I take it though that this can be seen as a mere notational issue, since such a dependency is clearly unavoidable. Second, and much more importantly, Hitchcock himself does not relativize his original definition of actual causation to a given model, but wants to account for causation as such that is independent of our description.

I maintain that such an understanding is misguided. However, I will postpone my argumentation for this to the next but one section of this chapter, where I will also discuss the philosophic implications of this differentiation between a pragmatic understanding of causation relativized to a given model by my definition (H AC) and a more metaphysical understanding of causation by Hitchcock's original formulation. Before that, I will illustrate how my reconstructions (H AR) and (H AC) can be practically applied in the analysis of example cases in the next section of this chapter.

Let me start my discussion with Hitchcock's original definition of counterfactual dependence in a solution of a causal model:[30]

Definition (Counterfactual Dependence in a Solution of a Causal Model, Hitchcock) **(H CDSCM)**

The value z of variable Z *depends counterfactually* on the value x of variable X in a solution $\boldsymbol{R}(\boldsymbol{V})$ of a causal model $\boldsymbol{M} := \langle \boldsymbol{V}, \boldsymbol{E} \rangle$ if and only if the following holds:

a) In the solution $\boldsymbol{R}(\boldsymbol{V})$ it is $X=x$ and $Z=z$.

b) There are values $x' \neq x$ and $z' \neq z$, such that replacing the equation Ex for X with the new equation $E'x := X=x'$ in $\boldsymbol{E}$ yields the result $Z=z'$ for variable Z.

Here, the notation of clause b) in (H CDSCM) is the shorthand of the following rather long condition: 'There are values $x' \neq x$ and $z' \neq z$, such that the following holds for the system $\boldsymbol{E}'$ that results from system $\boldsymbol{E}$, when we replace the equation Ex for X with the new equation $E'x := X=x'$ that renders X exogenous. If we enter the values $\boldsymbol{u}$ that the variables $\boldsymbol{U} \subseteq \boldsymbol{V}$ which are exogenous in $\boldsymbol{E}$ have in solution $\boldsymbol{R}(\boldsymbol{V})$ into the new system of structural equations $\boldsymbol{E}'$ (which also contains X as a new exogenous variable with value $X=x'$), this yields a new solution $\boldsymbol{R}'(\boldsymbol{V})$ in which the result for variable Z is $Z=z'$.'

Clearly, a surgical intervention is employed by this definition in order to find out whether a certain value z of variable Z depends counterfactually on the value x of another variable X in a solution of $\boldsymbol{M}$. For, observe that replacing the equation Ex for X with a new equation $E'x := X=x'$ in clause b) above amounts to surgically intervening in the model $\boldsymbol{M}$. This intervention

[30] Cf. [cH 01], p. 283 for Hitchcock's own original definition of counterfactual dependence in a solution of a causal model.

only locally changes the causal model, i.e. only the value of X is surgically altered, whereby only the values of descendant variables of X are affected. The resulting values of the descendant variables Y of X are then determined according to their corresponding structural equations E_Y that contain X or another descendant of X etc in their arguments. All other variables apart from the descendants of X keep their previous values.[31]

It would lead too far away to examine the philosophic significance of an intervention on a variable X in a causal model $\boldsymbol{M}$ here. I consider the concept of an intervention to be well-known in the literature about causation. For a deeper examination of the formal requirements that interventions have to fulfill and for a motivation how an intervention on a variable X in a causal model $\boldsymbol{M}$ can graphically be understood, I refer the reader to Woodward's illuminating discussion of this concept.[32]

For the purposes here it is sufficient to understand an intervention on variable X in model $\boldsymbol{M}$ as an ideal manipulation of the model that only affects variable X directly and that removes X from the influence of its ancestors. Since only the descendants of X are affected by an intervention and only the structural equation E_X of variable X is replaced, an intervention only minimally disturbs the causal model $\boldsymbol{M}$. Concerning the relation of an intervention to the situation $\boldsymbol{S}$ that is modeled by $\boldsymbol{M}$, the intervention on X as can be understood as a hypothetical manipulation of the event c in S that is formalized by the actual value x of X in $\boldsymbol{M}$. In particular, such a hypothetical manipulation of S is understood to be *ideal* in the sense that it is independent of human agency.

---

[31] In this context, I call Y a descendant of X iff X is a parent of Y or a parent of a parent of Y and so forth.

[32] Cf. [jW 03], pp. 327-336.

Let me add a word of clarification here, since Hitchcock did not formally distinguish this definition (H CDSCM) of counterfactual dependence in a solution of model M from his earlier definition (H CD) of counterfactual dependence for the set-up of the model M. In striking contrast to (H CDSCM), Hitchcock's definition (H CD) was first independent of the values of X and Z, because it existentially quantified over all possible values of all variables in V, and was second trivially independent of a solution in M, because the structural equations E in model M were just generated with the aid of (H CD).

The definition (H CDSCM) above decisively enters in Hitchcock's definition of an active route in a solution of a causal model.[33] In this context, a *route* between two variables X and Z in causal model $M := \langle V, E \rangle$ is an ordered sequence of variables $(X,Y_1,\ldots,Y_n,Z)$ in V such that each variable in the sequence is a parent of its successor in the sequence. And a variable Y is called *intermediate* between X and Z if and only if it belongs to some route between X and Z.

In the following, I will mainly discuss my reconstruction of Hitchcock's account. His original formulation neglects to mention the intrinsic dependence of the activity of a route on a causal model and its solution. However, when reading Hitchcock's original formulation, it is clear that this dependence is both intended and unavoidable.

My reconstruction of the definition of an active route is the following:

Definition (Active Route, Hitchcock) **(H AR)**

The route $(X,Y_1,\ldots,Y_n,Z)$ is *active in a solution* $R(V)$ *of the causal model* $M := \langle V, E \rangle$ if and only if the value z of Z depends counterfactually upon the value x of X within the resulting solution $R'(V)$ of the new system of equations E', constructed from E as follows:

[33] Cf. [cH 01], p. 286 for Hitchcock's original definition of an active route.

For all Y in V, if Y is intermediate between X and Z, but does not belong to the route $(X,Y_1,\ldots,Y_n,Z)$, then replace the equation E_Y for Y with a new equation $E'_Y := Y=y$ that sets Y constant to its value y in the solution $R(V)$.
If there are no intermediate variables that do not belong to this route, then E' is just E.[34]

Let me try to briefly convey the underlying idea of this definition. First, the route $(X,Y_1,\ldots,Y_n,Z)$ between variables X and Z in a solution of causal model M formalizes a chain of mechanisms at work in the situation S between the events c and e that are formalized by the values x of X and z of Z. Each variable Y_i in the route $(X,Y_1,\ldots,Y_n,Z)$ in question represents to a set of possible alterations of an event d_i in situation S. And to each of these sets of event alterations d_i a mechanism denoted by E_{Yi} corresponds that incorporates the local laws in situation S that determine d_i. In this way, the route $(X,Y_1,\ldots,Y_n,Z)$ corresponds to the concatenation of mechanisms denoted by $E_X \circ E_{Y1} \circ \ldots \circ E_{Yn} \circ E_Z$ that links events c and e in the situation S. The definition now follows the intuition that in order to analyze the particular effect of a change of event c along this route $(X,Y_1,\ldots,Y_n,Z)$ on event e, all other mechanisms apart from E_X, E_{Y1}, ..., E_{Yn}, and E_Z have to be screened off.
In this understanding, the route $(X,Y_1,\ldots,Y_n,Z)$ is active in the actual solution of model M if and only if a change of event c propagates exactly through the concatenation of mechanisms $E_X \circ E_{Y1} \circ \ldots \circ E_{Yn} \circ E_Z$, and no other mechanisms in situation S, and thereby changes the event e. Formally, the isolation of this chain of mechanisms $E_X \circ E_{Y1} \circ \ldots \circ E_{Yn} \circ E_Z$ linking c and e is achieved by freezing the other intermediate variables Y in the model M not belonging to the route $(X,Y_1,\ldots,Y_n,Z)$ by replacing their equations

[34] Observe that here $R'(V)$ is identical with the actual solution $R(V)$ of the model M, since the values y of the intermediate variables Y have not been changed by going over from E to E', they have just been frozen.

E_Y with new equations $E'_Y := Y=y$ that set Y constant to their actual values y.

In order to prevent possible confusions, let me state again that the above definition (H AR) and also its motivation are only my reconstructions of Hitchcock's view. His original formulation of an active route was the following:[35]

> **Definition (Active Route, Hitchcock's Misleading Original)** **(HMO AR)**
>
> The route $(X,Y_1,\ldots,Y_n,Z)$ is active in the causal model $\langle V,E\rangle$ if and only if Z depends counterfactually upon X within the new system of equations E', constructed from E as follows:
> for all Y in V, if Y is intermediate between X and Z, but does not belong to the route $(X,Y_1,\ldots,Y_n,Z)$, then replace the equation for Y with a new equation that sets Y equal to its actual value in E.
> (If there are no intermediate variables that do not belong to this route, then E' is just E.)

Apparently, Hitchcock's formulation neglects that the fact whether a route is active in a model M or not depends on the solution of a model. However, I take (HMO AR) just to be an elliptical formulation of (H AR), since this dependency on a solution of a model is inherited from the definition of counterfactual dependence in a solution of a causal model (H CDSCM). And in his original formulation of this definition (H CDSCM), Hitchcock did not suppress this dependency.[36]
Let me continue to describe Hitchcock's original formulation of his account of token causation. With this concept of the activity of a route in a

[35] Cf. again [cH 01], p. 286.
[36] Cf. again [cH 01], p. 283.

causal model *M*, Hitchcock defines actual causation as the following, in my view rather metaphysical, notion:[37]

Definition (Actual Causation, Hitchcock's Misleading Original) **(HMO AC)**

Let *c* and *e* be distinct occurring events, and let *X* and *Z* be variables, such that the values of *X* and *Z* represent alterations of *c* and *e* respectively. Then *c* is a *cause* of *e* if and only if there is an active causal route from *X* to *Z* in an *appropriate causal model* $\langle V, E \rangle$.

Apparently, this definition equates token causation in a situation *S* to the existence of an appropriate causal model *M* that contains an active route. With this, Hitchcock attempts to give an account of the problem what it means for an event to cause another event as such, independently of our representation of these events and their surrounding situation *S*. And there are several places in Hitchcock's original discussion, where he implicitly defends this goal.[38] The decisive notion in this context here is the one of an appropriate causal model. Admittedly, Hitchcock gives certain criteria for determining the appropriateness of a model.[39] However, I would like to postpone the discussion of this notion of appropriateness and also of its associated problems until the next but one section.

In the introduction of this thesis I already mentioned the distinction that I make between pragmatic accounts of causation relativized to a model and metaphysical accounts of causation as such. In my view, this formulation of definition (HMO AC) here is a clear attempt to account for causation as such. However, even a short synoptic discussion of this distinction would lead us too far away in this section. So, I ask the reader to bear with me for this and the next section, where I would first like to motivate how the

[37] Cf. [cH 01], p. 287.
[38] Cf. for example [cH 01], p. 274.
[39] Cf. [cH 01], p. 287.

leading idea of Hitchcock's account to analyze token causation can be understood and how a meaningful account of token causation relativized to a model can be condensed out of it.

Definition (HMO AC) equates token causation of event *e* by event *c* in a situation *S* to the existence of an active route in an appropriate causal model *M*. We have seen above that the route $(X,Y_1,\ldots,Y_n,Z)$ in a solution of a model *M* formalizes a chain of mechanisms $E_X \circ E_{Y1} \circ \ldots \circ E_{Yn} \circ E_Z$ at work in the situation *S* between the events *c* and *e* that are formalized by the values *x* of *X* and *z* of *Z* in model *M*. This route was active in the actual solution of the model if and only if a change of event *c* propagated exactly through the concatenation of mechanisms $E_X \circ E_{Y1} \circ \ldots \circ E_{Yn} \circ E_Z$, and no other mechanisms in situation *S*, and thereby changed the event *e*.
I would motivate this notion of an active route as a way to formally express our intuition of a *causal process* linking the events *c* and *e*. In this way, we would understand a causal process to be a concatenation of mechanisms $E_X \circ E_{Y1} \circ \ldots \circ E_{Yn} \circ E_Z$ at work in the given situation *S* linking events *c* and *e*, so that an alteration of event *c* would propagate exactly along this chain to change the event *e* with the other mechanisms in situation *S* not belonging to the causal chain in question being screened off. In my view, this idea of a causal process linking the events *c* and *e* is a very natural motivation of the role of an active route. However, Hitchcock in his original discussion does not speak of causal processes.

Let me now finally implement this rationale of an active route in a causal model as formal expression of a causal process into a definition of token causation relativized to a model. In my view, the straightforward reduction and also clarification of Hitchcock's original definition (HMO AC) given above is the following:

Definition (Actual Causation, Hitchcock) **(H AC)**

Let c and e be distinct occurring events in a situation S, and let $M := \langle V, E \rangle$ be a causal model formalizing S, such that the following holds: There are variables X and Z in V, such that the values x of X and z of Z in the actual solution $R(V)$ of $\langle V, E \rangle$ represent the actually occurring events c and e in S and such that their non-actual values x' and z' represent alterations of c and e respectively.
Then c is an *actual cause* of e *according to the model* M if and only if in the actual solution $R(V)$ of M there is an active route from X to Z.

The decisive change from Hitchcock's original formulation (HMO AC) to this reduction (H AC) intended to define token causation relativized to a model lies in the role of the causal model in these formulations. In (HMO AC), Hitchcock existentially quantifies over all appropriate models M in order to define what a cause in a situation S is – independently of our description of this situation. For, this is exactly what our causal models M are: They are descriptions of the given situation S. In essence, (HMO AC) rules that c is a cause of e in situation S if and only if there is an appropriate description M of S, according to which there is a causal process in S linking c and e. In contrast to this, (H AC) qualifies token causation in situation S to a fixed description M of this situation from the very beginning. (H AC) rules that c is a cause of e according to a fixed description M of situation S if and only if according to this fixed description M there is a causal process in S linking c and e.

Apart from this, the only formal difference between (HMO AC) and (H AC) is the following: Again, my formulation (H AC) explicitly mentions that the activity of a route depends on the solution of the model *M*. And again, this dependence is inherited from the definition of counterfactual dependence in a solution of a causal model (H CDSCM) and hence unavoidable.

As already mentioned, I will discuss the philosophic implications of this differentiation between a pragmatic understanding of causation relativized to a given model by my definition (H AC) and a more metaphysical understanding of causation by Hitchcock's original formulation (HMO AC) in the next but one section of this chapter. There, I will also argue more elaborately for the advantages of my pragmatic understanding.
In the next section, I will discuss two examples, in order to illustrate how my reconstruction (H AC) can be applied. I will employ these examples in my argumentation that my reconstruction of Hitchcock's account with definitions (H AR) and (H AC) seems to have a striking advantage in comparison to Lewis's classic counterfactual account regarding two of the main three problems of the latter – asymmetric over-determination and intransitivity of causation.

HITCHCOCK'S SUCCESSFUL ANALYSIS OF EXEMPLARY CASES OF ASYMMETRIC OVER-DETERMINATION AND OF THE INTRANSITIVITY OF CAUSATION

In this section, I will illustrate the functioning of my reconstruction of Hitchcock's definitions (H AR) and (H AC). Concretely, I will discuss how they are applied to analyze two of Hitchcock's favorite examples.

Hitchcock constructed the first of these examples to motivate that actual causation is not a transitive relation in general. I will state the main points of Hitchcock's discussion of this example and maintain that an analysis of this example with (H AR) and (H AC) is in accord with our causal intuitions about this example. Furthermore, I will illustrate that Lewis's classic counterfactual account fails in analyzing this example for obvious reasons.

The second example was developed by Hitchcock to give a case of asymmetric over-determination of an effect *e*, where the actual cause *c* pre-empts another event *d*, that would otherwise actually cause the effect *e*. Following Hitchcock, I will show that an analysis of this example with (H AR) and (H AC) will be in accord with what I take to be our causal intuitions about this example. Again, I will show that Lewis's classic counterfactual account cannot analyze this example successfully.

The discussion of these two exemplary cases then gives evidence to the conjecture that Hitchcock's account, reconstructed by (H AR) and (H AC) can surmount two of my featured three problems of the classic counterfactual account – pre-emption and the intransitivity of actual causation.

Let me start my discussion with the first example 'Boulder' that illustrates an intransitivity of actual causation. The cover story of 'Boulder' was the following:[40]

> **Boulder:** 'A boulder is dislodged, and begins rolling ominously towards a hiker. Before it reaches him, the hiker sees the boulder and ducks. The boulder sails harmlessly over his head with nary a centimeter to spare. The hiker survives his ordeal.'

As already mentioned, our pre-theoretic causal intuition about this situation was the following: The falling of the boulder causes the hiker to duck, and the ducking of the hiker causes his survival. However, the falling of the boulder does not cause the survival of the hiker, because the hiker would also have survived if the boulder had not rolled. In this way, this example shows a failure of composition in our pre-theoretic causal intuition. In the following, I demonstrate that the analysis by Hitchcock's account of actual causation is exactly in accord with this intuition of ours for this example.

Hitchcock's causal model for analyzing this example was the following: There were three bi-valued variables, *B*, *D* and *S*, with B being exogenous and D and S being endogenous. The meta-assignment of event-alterations to the values of these variables were:

$$B = \begin{cases} 0, & \text{if the boulder does not roll} \\ 1, & \text{if the boulder does roll} \end{cases}$$

$$D = \begin{cases} 0, & \text{if the hiker does not duck} \\ 1, & \text{if the hiker does duck} \end{cases}$$

$$S = \begin{cases} 0, & \text{if the hiker does not survive} \\ 1, & \text{if the hiker does survive} \end{cases}$$

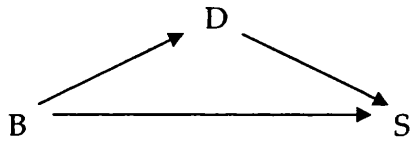
[40] Cf. again [cH 01], p. 276 for Hitchcock's own exposition of the example. For Hitchcock's own discussion and analysis of the 'Boulder' example, cf. pp. 295-299 of [cH 01].

The resulting structural equations for the endogenous variables *D* and *S* in Hitchcock's model were:

$$D = \begin{cases} 0, & \text{if B} = 0 \\ 1, & \text{if B} = 1 \end{cases} \quad \text{and} \quad S = \begin{cases} 0, & \text{if B} = 1 \text{ and D} = 0 \\ 1, & \text{if B} = 1 \text{ and D} = 1 \\ 1, & \text{if B} = 0 \text{ and D} = 0 \\ 1, & \text{if B} = 0 \text{ and D} = 1 \end{cases}$$

In the actual situation, *B* took the value 1, which yielded the actual solution *B = 1*, *D = 1*, and *S = 1*.

The graphical representation of this model of the 'Boulder' example finally was:



Verify first that *B = 1* is an actual cause of *D = 1* in the analysis by (H AC). In order to be precise, this is an elliptical formulation expressing that the event-alteration corresponding to *B = 1* is an actual cause of the event-alteration corresponding to *D = 1* in Hitchcock's analysis of the underlying situation, i.e. the rolling of the boulder causes the hiker to duck. I will use this identification of values of variables in a model with their associated events throughout this thesis for reasons of improved clarity.

In order to verify that *B = 1* is an actual cause of *D = 1* now, we have to show that the route (B,D) is active in the actual solution of our causal model according to (H AR). Observe that there are no intermediate variables between *B* and *D*, so that no structural equation is replaced in the model. Hence, for route (B,D) to be active now, the actual value *1* of variable *D* has to depend counterfactually on the actual value *1* of variable *B* in the actual solution of our model according to (H CDSCM). And this is the case, since in our causal model in our actual solution an intervention on variable *B* setting the value of *B* to *B = 0* yields the result *D = 0* for variable *D*. Here, I identify the replacing of the equation E_B for variable *B*

with the new equation $E'_B := B=0$ according to definition (H CDSCM) with setting of B to $B = 0$, in order to achieve a greater simplicity of notation again.

Analogously, verify that $D = 1$ is an actual cause of $S = 1$ in the analysis by (H AC) by showing that the route (D,S) is active in the actual solution of our causal model according to (H AR). For this, the actual value *1* of variable S has to depend counterfactually on the actual value *1* of variable D in the actual solution of our model according to (H CDSCM), since again there are no intermediate variables between D and S, so that no structural equation is replaced in the model. And this is again the case, since in our causal model in our actual solution with $B = 1$ an intervention on variable D setting $D = 0$ yields the result $S = 0$. With this, the route (D,S) is active, so that under Hitchcock's analysis also the ducking of the hiker causes him to survive.

Finally, verify that $B = 1$ is not an actual cause of $S = 1$ in the analysis by (H AC) by showing that neither the route (B,S) nor the route (B,D,S) is active in the actual solution of our causal model according to (H AR). Show first that route (B,D,S) is not active. Since there are no intermediate variables between B and S that do not belong to the route (B,D,S), this means that we have to show that the actual value *1* of variable S has does not depend counterfactually on the actual value *1* of variable B in the actual solution of our model according to (H CDSCM). And this is again the case, since in our causal model in our actual solution an intervention on variable B setting $B = 0$ yields the result $D = 0$, and this in turn yields $S = 0$. With this, the route (B,D,S) is not active according to (H AR). Show now that route (B,S) is not active. This time, variable D is intermediate between B and S and does not belong to the route (B,S). This means that we have to show that the actual value *1* of variable S does not depend counterfactually on the actual value *1* of variable B in the solution of the modification of our model, in which the equation E_D for D is replaced with

the new equation $E'_D := D{=}1$ that sets D constant to its actual value 1 in the actual solution of the model. And this is again the case, since in this modified causal model with variable D frozen at its actual value $D = 1$ an intervention on variable B setting $B = 0$ still yields the result $S = 1$. With this, the route (B,D) is not active, too, according to (H AR). Taken together, the rolling of the boulder does not cause the hiker to survive, since neither of the routes (B,S) and (B,D,S) is active.

With this, the verdict that an analysis by Hitchcock's account provides for this example is in accord with our causal intuition of this example: The falling of the boulder actually causes the hiker to duck, and the ducking of the hiker actually causes his survival in this situation. However, the falling of the boulder does not cause the survival of the hiker, because the hiker would also have survived, if the boulder had not rolled. In this way, Hitchcock's account of actual causation can successfully deal with this case containing an intransitivity of actual causation.

More importantly, one striking advantage of Hitchcock's account in comparison to Lewis's account becomes apparent in this example: For, an analysis of this example by Lewis's classic account featured in his article *'Causation'* delivers a verdict that contradicts our causal intuition. Since this account of actual causation has the assumption of transitivity built in, it rules in this example that the rolling of the boulder causes the survival of the hiker. This can easily be verified. Lewis's original account of token causation for events that actually occur can be summarized by the following two definitions:[41]

[41] Cf. [dL 73], pp. 198-200 for Lewis's original definition of counterfactual dependence.

Definition (Causal Dependence, Lewis) **(L CD)**

For two actually occurring events c and e, event e is *causally dependent* on event c if and only if the following counterfactual holds:
If c had not occurred, then e would not have occurred.

Definition (Actual Causation, Lewis) **(L AC)**

For two actually occurring events c and e, event c *actually causes* event e if and only if the following holds:
There is a chain of events $d_1,\ldots,d_n$, such that d_1 is causally dependent on c, d_2 is causally dependent on d_1, ... , and e is causally dependent on d_n.

Lewis gives certain criteria under which a counterfactual as mentioned in (L CD) is true. For this reason, Lewis gives an elaborate semantics of so-called possible worlds in order to account for the truth of counterfactuals.[42] However, it is beyond the scope of this thesis to describe this semantics in detail. For our discussion in this chapter, and in the rest of the thesis, it is absolutely sufficient to take definition (L CD) at face value and to assume a naïve pre-theoretic understanding of the truth of a counterfactual.

There is just one condition that has to be followed with regard to this definition (L CD). Lewis calls this condition *non-backtracking*.[43] In essence, non-backtracking requires that in evaluating the featured counterfactual 'If c had not occurred, then e would not have occurred.' in definition (L CD), the other events g in the situation S that cause event c but not vice versa are held fixed. The main idea behind this requirement is to prevent

[42] Cf. [dL 73], pp. 196-198 for Lewis's own discussion of criteria governing the truth of counterfactuals in this context. For a more elaborate discussion of Lewis's possible world semantics, cf. his book *Counterfactuals*, [dL 73b].

[43] Cf. mainly the elaborate discussion of the non-backtracking criterion in [dL 79]. Cf. also p. 275 of [cH 01].

the account from incorrectly ruling that g is also causally dependent on c. The functioning of this criterion will become clearer in the discussion of the following three examples featured in this chapter. For further discussion of the non-backtracking criterion, I would like to refer the reader to Lewis's original text.[44]

The immediate observation that can be made form these two definitions is that Lewis's builds a transitivity of causation into his account. This becomes obvious by realizing that for two events c and e, if event e is causally dependent on event c according to definition (L CD), then c already causes e. For, the two events c and e form a trivial causal chain according to definition (L AC), consisting only of these two events. This is the stipulation of transitivity in Lewis's classic account of causation, against which Hitchcock argues. And this example 'Boulder' delivers strong evidence for this case, as will see now:

Consider, how Lewis's account analyzes this example: Let b_1 denote the event that the boulder rolls, corresponding to $B = 1$ in Hitchcock's causal model. Analogously, let d_1 and s_1 denote the events that the hiker ducks and that the hiker survives respectively, corresponding to $D = 1$ and $S = 1$ respectively in Hitchcock's causal model.

Observe that first the event d_1 causally depends on the event b_1 according to (L CD), if we take this definition at face value. For, if the boulder had not rolled, the hiker would not have ducked in the situation. Observe second that also the event s_1 causally depends on the event d_1 according to (L CD), such that the survival of the hiker is causally dependent on the ducking of the hiker in Lewis's analysis.

Here, the non-backtracking criterion enters. When we consider the situation in which the hiker did not duck, we hold the rolling of the boulder fixed. Two things are achieved by this: First, we prevent b_1 from being incorrectly analyzed as causally dependent on d_1. Second, we limit

[44] Cf. again [dL 79].

the disturbance of the situation to a minimum. The interesting consequence of this minimal disturbance is that the event s_1 changes as a result of not letting the change in event d_1 track back to b_1.

Both these observations are firmly supported not only by our causal intuition, but also by our pre-theoretic understanding of counterfactuals: If the boulder had not rolled, the hiker would not have ducked in our example; and if the hiker had not ducked in this situation with the boulder rolling, he would not have survived.

However, the result that an application of Lewis's definition of actual causation (L AC) provides stands in contradiction our causal intuition. For, if d_1 causally depends on b_1 and if s_1 causally depends on d_1, then there is a trivial chain b_1,d_1,s_1 of causally depending events, such that according to (L AC) b_1 actually causes s_1. So, the verdict of an application of Lewis's definition (L AC) to this case is that the rolling of the boulder causes the survival of the hiker, strongly contradicting our causal intuition about this case.

Lewis's only way out of this problem would be to deny that d_1 causally depends on b_1 and that s_1 causally depends on d_1. He could for example argue that the survival of the hiker is hardly an event in our usual understanding, much more a fact that is dependent on its description. In this line of argument, Lewis could limit his account to events in our natural understanding. However, I regard this way out as not being satisfactory. For, I take it that we have a strong intuition in this case that the ducking of the hiker actually causes his survival in the given circumstances with the boulder rolling. And an account of actual causation that cannot account for this intuition of ours falls short in its intended goal in my opinion.

On the first glance, another possible line of argument for Lewis to deny that d_1 causally depends on b_1 and that s_1 causally depends on d_1 could also be that our given causal model is not appropriate for the example

situation. However, in my view, such a line of argument would be on a par with the one above discrediting the survival of the hiker as not being a proper event. For, apart from extracting the events b_1, d_1 and s_1 and their alterations, the only other ingredient for the generation of a causal model is the complete counterfactual information that is stipulated about the situation in the modeling procedure. And I take this counterfactual information to be a formal reflection of our intuition about this example. I maintain that we are interested in these events or facts b_1, d_1 and s_1 and their alterations in the way we describe them. And we are not only interested, but we also have clear intuitions about them. In my view, not being able to account for these intuitions is not a satisfactory result when it comes to formalizing causation.

Let me now deal with the second of Hitchcock's favorite examples that contains a case of asymmetric over-determination in which one event causally pre-empts another one. This is the example 'Backup' with the following cover story:[45]

> **Backup:** 'An assassin-in-training is on his first mission. Trainee is an excellent shot: if he shoots his gun, the bullet will fell the victim. A supervisor is also present, in case the trainee has a last minute loss of nerve and fails to pull the trigger. If the trainee does not shoot, the supervisor will shoot the victim herself. In fact, the trainee performs admirably, firing his gun and killing the victim.'

Hitchcock assumes the following assignment of values of variables to event-alterations:

---

[45] Cf. [cH 01], p. 276 for the cover story and pp. 287-288 for Hitchcock's original discussion of this example.

$$T = \begin{cases} 0, & \text{if the trainee does not shoot} \\ 1, & \text{if the trainee does shoot} \end{cases}$$

$$S = \begin{cases} 0, & \text{if the supervisor does not shoot} \\ 1, & \text{if the supervisor does shoot} \end{cases}$$

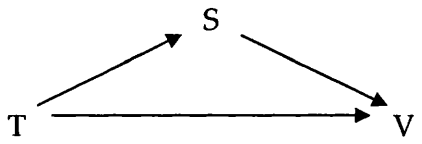
$$V = \begin{cases} 0, & \text{if the victim does not die} \\ 1, & \text{if the victim does die} \end{cases},$$

With this, the resulting structural equations in Hitchcock's causal model are the following:

$$S = \begin{cases} 0, & \text{if T} = 1 \\ 1, & \text{if T} = 0 \end{cases} \quad \text{and} \quad V = \begin{cases} 0, & \text{if T} = 0 \text{ and S} = 0 \\ 1, & \text{if T} = 0 \text{ and S} = 1 \\ 1, & \text{if T} = 1 \text{ and S} = 0 \\ 1, & \text{if T} = 1 \text{ and S} = 1 \end{cases},$$

yielding the actual solution $T = 1$, $S = 0$ and $V = 1$.

The graphical representation of this model of Hitchcock's is finally:



In my opinion, our causal intuition in this case is rather unambiguous: In our understanding, the shooting of the trainee actually causes the death of the victim and pre-empts the shooting of the supervisor that would have otherwise killed the victim. In this way, we have a clear case of asymmetric over-determination here, in which the actual cause, namely the shooting of the trainee, pre-empts another event, namely the shooting of the supervisor, from actually causing the result, i.e. the death of the victim.

Let me verify now that the verdict that an analysis by Hitchcock's account of actual causation provides for this example is again in accord with our causal intuition. I demonstrate that in an application of my reconstructed definition (H AC) to the causal model given above the shooting of the

trainee, formalized by $T = 1$, is an actual cause of the death of the victim, formalized by $V = 1$. For this purpose, route (T,V) has to be active according to my reconstructed (H AR) in the actual solution $T = 1$, $S = 0$ and $V = 1$ of our causal model. Apparently, variable S is intermediate between T and V and does not belong to the route (T,V). This means that we have to show that the actual value *1* of variable V does depend counterfactually on the actual value *1* of variable T in the solution of the modification of our model, in which the equation E_S for S is replaced with the new equation $E'_S := S{=}0$ that sets S constant to its actual value 0. And this is again the case, since in this modified causal model with variable S frozen at its actual value $S = 0$ an intervention on variable T setting $T = 0$ yields the result $S = 0$, in this way changing the value of V. Hence, the route (T,V) is active according to (H AR), and $T = 1$ is an actual cause of $V = 1$ according to (H AC), so that Hitchcock's analysis is again in accord with our causal intuition in this example.

To explore Hitchcock's analysis a little further, let me also verify that under Hitchcock's analysis the shooting of the supervisor would have actually caused the death of the victim, if the trainee had not shot. With this, an analysis by Hitchcock's account of actual causation would again be in accord with our causal intuition by being able to work out the pre-emption structure in this example. I demonstrate that in an application of definition (H AC) to the causal model given above the shooting of the supervisor, formalized by $S = 1$, would be an actual cause of the death of the victim, formalized by $V = 1$, under the changed circumstances in which the trainee did not shoot, formalized by $T = 0$. For this purpose, I have to show that in the new solution $T = 0$, $S = 1$ and $V = 1$ of the causal model, resulting from changing the value of the exogenous variable to $T = 0$ and letting this information propagate in the model, the route (S,V) is active according to (H AR). For this, the actual value *1* of variable V has to depend counterfactually on the actual value *1* of variable S in this changed

solution of our model. And this is again the case, since in this actual with $T = 0$ an intervention on variable S setting $S = 0$ yields the result $V = 0$. As a result, the route (S,V) is active in this changed solution, so that under Hitchcock's analysis the shooting of the supervisor would have caused the death of the victim, if the shooting of the trainee had not pre-empted this.

Consider now, how Lewis's account analyzes Hitchcock's example 'Backup': Let t_1 denote the event that the trainee shoots, corresponding to $T = 1$ in Hitchcock's causal model. Analogously, let s_1 and v_1 denote the events that the supervisor shoots and that the victim dies respectively, corresponding to $S = 1$ and $V = 1$ respectively in Hitchcock's causal model.

Observe that first the event v_1 does not directly causally depend on the event t_1 according to (L CD). For, if the trainee had not shot, the supervisor would have shot, and the victim would still have died in the situation. So, for Lewis's account to rule that the shooting of the trainee is an actual cause of the death of the victim, we have to find a causal chain between t_1 and v_1. The only possible candidate for an intermediate link in a chain between t_1 and v_1 in this situation is s_1. However, v_1 is not causally dependent on s_1 in the actual situation. For, if the supervisor had shot in addition to the trainee, the victim would have been at least as dead as before. As a result, the events t_1, s_1, and v_1 do not form a causal chain according to (L AC).

For Lewis's account to work in this example, another event would have to be introduced into this situation to form a causal chain between t_1 and v_1. In his discussion of Lewis's account applied to this example Hitchcock pointed out that the standard attempt to rescue Lewis's account here would be to introduce the intermediate event of a bullet en route from the trainee's rifle to the victim.[46] Admittedly, the artificial introduction of this

[46] Cf. [cH 01], p. 276 and pp. 287-289.

event r_1 would allow Lewis's account to get around this problem: Event r_1 would causally depend on t_1, and v_1 would causally depend on r_1.[47]

However, there are two problems in my view with this ad hoc addition of event r_1. First, the ad hoc introduction of this new event amounts to maintaining that our given causal model is not appropriate for the example situation However, with the same argumentation as I brought forward against the second possible rescue strategy of Lewis's account for the previous 'Boulder' example, I maintain this model is very well justified. We extracted the events, or facts, t_1, s_1 and v_1 out of the description of the situation according to the questions that we had about this situation. And again, we were not only interested in these facts, but we also had clear intuitions about them. The introduction of this new event r_1 does not really answer these questions of ours that led to the construction of our causal model, but avoids them.

Second, the core idea of this rescue attempt is to take advantage of the stipulated transitivity in Lewis's definition of actual causation (L AC). But the previous example 'Boulder' delivered a strong argument against such a general stipulation.

Hitchcock's account can analyze this example 'Backup' in accord with our causal intuition without having to assume a transitivity of causation. And this is the second advantage of Hitchcock's account that I wanted to highlight.

[47] For a verification of this, cf. again [cH 01], pp. 287-289.

HITCHCOCK'S NOTION OF APPROPRIATENESS OF A CAUSAL MODEL AND HIS CLAIM TO REDUCE CAUSATION TO COUNTERFACTUALS

I mentioned in the introduction of this thesis that token causal modeling can be understood to split the problem of analyzing token causation into two sub-problems: the first being the generation of a suitable causal model M to analyze a situation S, and the second is the definition of token causation according to this given model M.

In this section, I will argue that this differentiation into two sub-problems has a great advantage over classic metaphysical theories of causation: In qualifying causation to a given causal model M in the second sub-problem, our formal analysis can account for the fact that there is a multitude of possible descriptions of this situation that may all be equally justified, depending on which kinds of events are of interest in S. For, I will argue that the individuation of events and their alterations heavily depends on the questions that we would like our analysis of the situation S to answer. And very often we are not merely interested in causal relationships between events in a physical sense, but also between facts of the matter that are dependent on our descriptions of them.

Of course, an account of token causal modeling cannot be complete without clarifying how the first sub-problem is solved, i.e. how such a causal model M is generated. I tried to reconstruct Hitchcock's modeling procedure with steps M1) and M2) in the first section of this chapter. In this section I will examine Hitchcock's notion of the appropriateness of a causal model. I will argue that his original formulation of appropriateness (HMO AM) has an underlying metaphysical spirit, but that it can

successfully be reduced to a requirement of non-backtracking (H NB) that is consistent with my steps M1) and M2). I will summarize my reconstruction of Hitchcock's answer to the problem of model generation in the following way: Hitchcock has to start out with a certain description of a situation *S* which specifies the events and their alterations in which we are interested. This description can be represented as a basic set of super-exhaustive counterfactuals that has the non-backtracking property. This set of super-exhaustive counterfactuals uniquely determines a causal model ***M*** in such a way that the model rewrites the counterfactual information in this basic set in a more condensed way.

I will conclude this section with a short discussion of another remarkable tendency of Hitchcock's original account, namely to conceptually reduce causation to counterfactuals. Hitchcock attempts to express the central concept in his account, i.e. the activity of a route, with the aid of a special kind of counterfactual, a so-called *Explicitly Non-Foretracking* (ENF) *counterfactual*. I will maintain that this reduction falls short, since the concept of a surgical intervention has to enter in definition (H AR) and Hitchcock does not give any clues how an intervention could be expressed by counterfactuals.

In my opinion, the decisive notion in causal modeling is the one of a causal model. I do not regard this as a platitude however. The introduction of a causal model for me marks the transition from a metaphysical theory of causation that attempts to account for 'causation as such' to a more pragmatic modeling account with the more modest claim to analyze causal relationships according to a given model.

Lewis's classic account of token causation attempted to clarify under which conditions a certain event c causes another event e. These events were usual physical events in a naïve scientific understanding, or to formulate it crudely, 'events in the real world'.[48] In particular, c and e were understood to be independent of the descriptions employed for their individuation.

Lewis did not stop with this bold claim to account for 'causation as such'. Instead, he attempted to conceptually reduce the notion of causation to counterfactuals. This reductive idea becomes obvious in my reconstructions of Lewis's definitions of causal dependence (L CD) and actual causation (L AC). According to (L CD), the actually occurring event d_2 is causally dependent on the actually occurring event d_1 if and only if the following counterfactual holds: 'If d_1 had not occurred, then d_2 would not have occurred.'. Definition (L AC) then rules that for two actually occurring events c and e, event c actually causes event e if and only if there is chain of causally dependent events $d_1,\ldots,d_n$ linking c and e.

In this way, the auxiliary concept of causal dependence is expressed as a counterfactual relationship for Lewis. And with this, actual causation indirectly reduces to a counterfactual relationship, too, in Lewis's framework. Finally, Lewis defines conditions for the truth of these counterfactuals via a semantics of possible worlds employing a naïve underlying notion of truth in a possible world.[49] However, I do not want to go into the details of this attempt of Lewis to express token causation in purely acausal terms. I just wanted to emphasize the metaphysic nature of Lewis's project not only to account for token causation independently of our description but also to conceptually reduce causation to counterfactual relations.

---

[48] Cf. [dL 73] and also [dL 73b].
[49] Cf. [dL 73b].

In stark contrast to this, the introduction of a causal model in my opinion allows a clear formal analysis of various causal notions by a given model with a well-defined ontology. In the following, let me briefly mention the advantages that the introduction of a causal model together with a clear ontology delivers for the analysis of causal relations on the token level in my view.

The starting point for the construction of a structural causal model is a description of a given situation *S*, consisting of various singular events. The distinctive feature of a structural theory of modeling token causation is that it exploits counterfactual information about the occurring events in *S* for the analysis of their causal relationships. As a result, the first decisive problem in the introduction of a structural causal model for the analysis of token causation is what we admit as an event and its possible alterations. I argue that this individuation of events and their alterations heavily depends on the questions that we would like our analysis of the situation *S* to answer.

Consider the 'Boulder' example again: In analyzing this example our interest was in the question whether the rolling of the boulder is an actual cause for the survival of the hiker. The survival of the hiker is though hardly an event, even in the most liberal understanding of events. In classic philosophical terminology, the survival or not-survival of the hiker are classified as facts of the matter. So, if we want to answer whether the rolling of the boulder caused the hiker to survive, we have no other choice than to treat the survival of the hiker as a quasi-event with its possible alteration, the not-survival of the hiker. Questions like this concerning causal dependencies between facts that are individuated according to our interests in and descriptions of certain quasi-situations, rather than between physical events, are in no way far-fetched. On the contrary, in my view they are the norm in ordinary causal discourse.

Nevertheless, a treatment of these questions poses a problem for Lewis's account of causation, in which events and their alterations were understood as being strictly physical.[50] And problems do not stop here. Kim put forward a popular criticism of Lewis's classical counterfactual account in which he showed that there are types of quasi-situations containing quasi-events between which there are counterfactual dependencies that are not causal.[51] These quasi-events could be counterfactually dependent in virtue of the meaning of their descriptions, because they are logically or mathematically related, or because of certain non-causal laws or boundary conditions.[52]

This is what I regard as the main advantage of limiting causal analysis to a model: A model brings with it a well-defined ontology, and this ontology can contain exactly the quasi-events or facts and their alterations in which we are interested. In this way, the events which we analyze with the aid of this model do not have to be physical events and do not even have to be independent of our description of them. They can be abstract entities like facts that are individuated according to our descriptions.

In this way, apart from the three main problems of Lewis's classic counterfactual account that I stated in the introduction of this thesis – symmetric over-determination, asymmetric over-determination and stipulation of transitivity – I would like to count this as a fourth problem: Inability to account for a wide array of events and facts. Furthermore, let me also finally remark that definitions (L CD) and (L AC) can only deal with the occurrence or non-occurrence of events and not with more complicated alterations of events as can be done in a causal model containing multi-valued variables.

---

[50] Cf. again [dL 73] and [dL 73b].
[51] Cf. [jK 73].
[52] Cf. [nC 89].

Remarkably, despite this in my view major advantage of a restricted understanding of causal modeling to relativize causal relationships to given models, Hitchcock himself still champions the metaphysical aim to account for causation as such, independent of our description. For, his original definition of actual causation (HMO AC) ruled that *c* is an actual cause of *e* if and only if there is an active causal route from *X* to *Z* in an *appropriate causal model* $\langle V, E \rangle$.[53] In the following I will hence briefly discuss his original notion of appropriateness of a causal model. According to Hitchcock, there are at least the following three requirements for a model $M := \langle V, E \rangle$ to appropriately formalize a given situation *S*:[54]

Partial Characterization (Appropriateness of a Model, Hitchcock's Misleading Original) **(HMO AM)**

A causal model $M := \langle V, E \rangle$ is appropriate, only if the following holds:

i) The equations in *E* entail no false counterfactuals.

ii) The equations in *E* do not represent counterfactual dependence relations between events that are not distinct.

iii) *V* does not contain variables whose values correspond to possibilities that we consider to be too remote.

Only the third condition iii) mentions our considerations in describing the underlying situation *S*. Condition i) is formulated in a way that suggests that the set of counterfactuals, which I call the basic set of super-exhaustive counterfactuals, is independent of our description of the situation *S*. And condition ii) seems to imply that also the individuation of events is independent of our interest in situation *S*. In my view, this partial characterization is not only misguided but, with the exception of condition i), also rather empty.

---

[53] Cf. [cH 01], p. 287.
[54] Cf. again [cH 01], p. 287.

Let me briefly try to motivate this claim and clarify how (HMO AM) can be embedded into my reconstruction of Hitchcock's modeling account. The third condition iii) reduces to the requirement that we should individuate events and their alterations in *S* in a way that they are relevant for our analysis. If this requirement is not further specified, it is already contained in my description of the first step in modeling a situation, denoted by M1) in the first section of this chapter. Condition ii) can be reconstructed similarly as requiring that events and their alterations in *S* also have to be individuated in a way that disjoint variables can only formalize disjoint sets of event alterations. And such a criterion can be considered to be rather obvious.

Only condition i) of (HMO AM) is more interesting and concerns the second step M2) of my reconstruction of Hitchcock's procedure of modeling a situation in the first section of this chapter. Since Hitchcock's structural equations *E* encode the counterfactual information that is contained in the set of exhaustive counterfactuals describing situation *S*, condition i) reduces to a requirement of this set of counterfactuals. This requirement can be reconstructed as demanding that this set may not contain certain counterfactuals which we consider not to be admissible.

Hitchcock is rather vague about which counterfactuals should be excluded in this way. There is just one remark of his stating that the counterfactuals which are encoded by his structural equations *E* should not *backtrack*.[55] Unfortunately, Hitchcock does not clarify what such a non-backtracking requirement really means. In my opinion, it can be reconstructed in analogy to Lewis's non-backtracking criterion which I briefly mentioned in the last section of this chapter. In this way, I suggest the following reconstruction of Hitchcock's non-backtracking requirement:

[55] Cf. [cH 01], p. 280.

Constraint (Non-Backtracking, Hitchcock) **(H NB)**

The basic set of super-exhaustive counterfactuals constituting the causal model $M := \langle V, E \rangle$ describing a situation S has to be comprised in a way so that the following holds:
For all variables Y and W in set V, if W is counterfactually dependent on Y according to definition (H CD), then Y is not also counterfactually dependent on W.

This way, it is excluded that a change of the value w of variable W in the model automatically tracks back to the parents Y of W, if we test whether w might be an actual cause of another value z of a variable Z in model M. In order to illustrate this criterion (H NB), consider again the 'Boulder' example. In the last section of this chapter, I demonstrated that in our causal model with the actual solution B=1, D=1, and S=1, an intervention on variable D setting $D = 0$ yielded the result $S = 0$. This was because the change in the value of D to $D = 0$ did not track back to its parent variable B, for otherwise the value of B would have changed to $B = 0$, too, so that the route (D,S) would no longer have been active.
This is the only requirement that I would like to keep from Hitchcock's original notion of the appropriateness of a causal model expressed in (HMO AM): In Hitchcock's causal models M defined by (H CM), backtracking is excluded by requiring that the condition in (H NB) is fulfilled in the underlying set of super-exhaustive counterfactuals that constitutes M. Apparently, Hitchcock took this non-backtracking condition requiring that no variable Y can be both parent and child of another variable W very seriously. For, he repeatedly emphasized that his structural equations are *asymmetric* in this way.[56]

[56] Cf. for example [cH 01], pp. 275 and 280.

Summarizing, this brief argumentation has shown that Hitchcock's original notion of the appropriateness of a causal model (HMO AM) certainly has an underlying metaphysical spirit, but that it can successfully be reduced to a requirement that is consistent with my reconstruction of the modeling procedure with steps M1) and M2) in the first section of this chapter. I hope that I have given enough reason and motivation for the pragmatic view that M1) and M2) express: In qualifying causal analysis of a situation *S* to a model, we acknowledge that this analysis is driven by our interests in the situation and that the way we individuate events and quasi-events in *S* reflects these interests.

I mentioned in the introduction of this thesis that token causal modeling splits the problem of analyzing token causation into two sub-problems: the first being the generation of a suitable causal model ***M*** to analyze a situation *S*, and the second is the definition of token causation according to this given model ***M***. I argue that this differentiation into two sub-problems has the great advantage over classic metaphysical theories of causation of being able to account for the dependency of the causal analysis of a situation *S* on our interests in this situation. In qualifying causation to a given causal model ***M*** in the second sub-problem, our formal analysis can account for the fact that there is a multitude of possible descriptions ***M*** of this situation that may all be equally justified, depending on which kinds of events are of interest in *S*.

Of course, an account of token causal modeling cannot be complete without clarifying how the first sub-problem is solved, i.e. how such a causal model ***M*** is generated. I tried to reconstruct Hitchcock's modeling procedure with steps M1) and M2) in the first section of this chapter. The discussion of Hitchcock's notion of appropriateness of a causal model in this section has added my definition of the non-backtracking criterion (H NB) to this procedure. In this way, I summarize my reconstruction of Hitchcock's answer to the problem of model generation in the following

way: Hitchcock starts out with a certain description of a situation *S* which specifies the events and their alterations in which we are interested. This description can be represented as a basic set of super-exhaustive counterfactuals that has the non-backtracking property. This set of super-exhaustive counterfactuals uniquely determines a causal model *M* in such a way that the model rewrites the counterfactual information in this basic set in a more condensed way. In this way, I regard Hitchcock's answer to the problem of model generation to be rather empty and to be more an account of model stipulation. However, this thesis is mainly concerned with the second problem of causal modeling, i.e. the definition of causation according to a given model. In this way, I settle for this result and do not attempt to clarify how a description of a situation can be reconstructed by a basic set of super-exhaustive counterfactuals.

Let me conclude this section with a short discussion of another remarkable tendency of Hitchcock's original account, namely to conceptually reduce causation to acausal terms. Hitchcock explicitly claims that if Lewis offered a reductive analysis of causation in purely acausal terms with the aid of counterfactuals, then his own account offers a reductive analysis of causation with the aid of active routes.[57] To be more precise, Hitchcock attempts to express the central concept in his account, i.e. the activity of a route, with the aid of a special kind of counterfactual, a so-called *Explicitly Non-Foretracking* (ENF) *counterfactual*.[58] In my understanding, Hitchcock's motivation for this is to conceptually reduce his notion of actual causation to purely counterfactual and hence acausal notions.

[57] Cf. [cH 01], p. 274.
[58] Cf. [cH 01], pp. 285-287 for Hitchcock's own comments about ENF counterfactuals and their correspondence to the notion of an active route.

Reconstructing Hitchcock's view in my terminology, a route $(Y,X_1,\ldots,X_n,W)$ is active in a solution $\boldsymbol{R}(\boldsymbol{V})$ of the causal model $\boldsymbol{M} := \langle \boldsymbol{V}, \boldsymbol{E} \rangle$ according to definition (H AR) if and only if there are two true counterfactuals

I1) $Y=y \;\&\; Z_1=z_1 \;\&\ldots\&\; Z_k=z_k \;\boxed{\rightarrow}\; W=w$ and

I2) $Y=y' \;\&\; Z_1=z_1 \;\&\ldots\&\; Z_k=z_k \;\boxed{\rightarrow}\; W=w'$

such that the following holds: In the counterfactual I1), y and w are the actual values of variables Y and W in the solution $\boldsymbol{R}(\boldsymbol{V})$, and $Z_1,\ldots,Z_k$ are exactly the variables in $\boldsymbol{V}$ that are descendants of Y and ancestors of W, but do not belong to the route $(Y,X_1,\ldots,X_n,W)$ in question, and $z_1,\ldots,z_k$ are their respective actual values in the solution $\boldsymbol{R}(\boldsymbol{V})$. In the counterfactual I2), $y' \neq y$ and $w' \neq w$ are non-actual values of variables Y and W.

Indeed, the truth of these counterfactuals under these circumstances is equivalent to the conditions required in definition (H AR) for the route $(Y,X_1,\ldots,X_n,W)$ to be active. For having $Z_1=z_1 \;\&\ldots\&\; Z_k=z_k$ in the antecedents of these counterfactuals corresponds to freezing the intermediate variables between Y and W at their actual values, in this way constructing a new system of structural equations $\boldsymbol{E}'$ as required by definition (H AR). The truth of both the first and the second counterfactual then literally expresses the counterfactual dependence of $W=w$ on $Y=y$ in this new solution $\boldsymbol{R}'(\boldsymbol{V})$ of $\boldsymbol{E}'$, again as required by definition (H AR). In order to briefly motivate calling these counterfactuals *explicitly non-foretracking*, observe that the freezing of variables $Z_1,\ldots,Z_k$ at their actual values $z_1,\ldots,z_k$ exactly achieves that the change in the value of Y to $y' \neq y$ cannot propagate or *foretrack* along other routes between Y and W than the one specified $(Y,X_1,\ldots,X_n,W)$.

With this, Hitchcock wants to express the activity of a route in purely counterfactual terms and hence reduce his notion of actual causation to purely counterfactual and hence acausal terms. However, I maintain that the reduction aimed for falls short. For, in my view the essence of

definition (H AR) not only lies in the way, in which a change of the value of Y propagates in the causal model, but also in the notion of a change in the model itself. For, both for freezing the intermediate variables $Z_1,\ldots,Z_k$ at their actual values $z_1,\ldots,z_k$ and for changing the actual value of Y into $y' \neq y$ we have to *intervene* in the causal model M. In my understanding, the concept of a surgical intervention is clearly a causal notion, and Hitchcock does not give any clues how an intervention could be expressed with the aid of a counterfactual. Hence, the mere equivalence of the truth of the two counterfactuals I1) and I2) above to the condition expressed in definition (H AR) for a route to be active is not enough for reducing (H AR) to counterfactual notions. Instead, we would also have to explain in counterfactual terminology how we can intervene on the variables in the antecedents of these counterfactuals, and this is where Hitchcock falls short.

The problem of keeping the notion of an intervention as an irreducible causal concept in an account of causation, like my reconstruction of Hitchcock's account with definitions (H AR) and (H AC), is rather profound and the debate about this is rather extensive. For this reason, I do not go deeper into this problem here and refer to Woodward's illuminating discussion of the notion of an intervention.[59]

[59] Cf. [jW 03], pp. 94-107.

HITCHCOCK'S DEFINITION OF CONTRIBUTORY CAUSATION EMPLOYING WEAKLY ACTIVE ROUTES FOR ANALYZING CASES OF SYMMETRIC OVER-DETERMINATION

In this section, I will deal with Hitchcock's account of contributory causation for cases of symmetric over-determination. Admittedly, the term *contributory causation* is my terminology, and it is borrowed from Pearl.[60] Hitchcock does not give this account a special name, he just states that his previous definition of actual causation, reconstructed by my (H AC), is not capable of analyzing cases appropriately in which an effect has two or more causes that work in parallel. For cases of this form, Hitchcock introduces his definition of a weakly active route, which I will reconstruct as (H WAR), and his definition of token causation for these cases, which I dub contributory causation and which I reconstruct as (H CC).

In my reconstruction of Hitchcock, an event c is then a contributory cause of another event e according to a certain model if and only if there is a weakly active causal route between the variables formalizing these events in this causal model. In this account of contributory causation with my reconstructed definitions (H WAR) and (H CC), Hitchcock deviates from his previous rationale of equating a causal process with an active route in a model. Now the formalization of a causal process between a contributory cause c and its effect e in a situation S is a weakly active route in a causal model M between the variables X and Z whose actual values x and z formalize these events c and e.

---

[60] Cf. my discussion of Pearl's account of modelling token causation in the next chapter of this thesis.

Still, a chain of mechanisms linking *c* and *e* in situation *S* corresponds to this route. However, in the criterion for this route to be weakly active this chain of mechanisms is no longer isolated, so that a change in event *c* can propagate along other mechanisms outside the chain to event *e*.

Again, Hitchcock himself does not relativize his original account of token causation for cases of symmetric over-determination to a given model, but wants to account for causation as such in these cases. However, referring to my argumentation before, I will continue to maintain that such an understanding is misguided.

As a point of application and comparison, I will show how my reconstructions of Hitchcock's definitions (H WAR) and (H CC) are applied in an analysis of Lewis's classic example of symmetric over-determination, the Firing Squad. Lewis famously claimed to have no clear intuition in this case. Many other people have one, though, and I will demonstrate that in the analysis of this example Lewis's account fails to analyze two events as contributory causes that are contributory causes according to our pre-theoretic causal understanding. In striking contrast to this, my reconstruction of Hitchcock's account with definition (H WAR) and (H CC) successfully detects these two contributory causes.

Let me start my discussion with my reconstruction of Hitchcock's definition of a weakly active route, which is the following:[61]

Definition (Weakly Active Route, Hitchcock) **(H WAR)**

The route $(X,Y_1,\ldots,Y_n,Z)$ is *weakly active in a solution* $R(V)$ *of the causal model* $M := \langle V, E\rangle$ if and only if there exists a set (possibly empty) of variables $\{W_1,\ldots,W_m\}$ in $V \setminus \{X,Y_1,\ldots,Y_n,Z\}$, and values

[61] Cf. [cH 01], pp. 289-290 for Hitchcock's original discussion of cases of symmetric over-determination.

$w_1,\ldots,w_m$ that lie within the redundancy range of $\{W_1,\ldots,W_m\}$ for this route and its values in $R(V)$, such that the following holds:
The value z of Z depends counterfactually upon the value x of X within the resulting solution $R'(V)$ of the new system of equations E', constructed from E as follows:
For each W_i, replace the equation E_{Wi} for W_i with the new equation $E'_{Wi} := W_i=w_i$ that sets W_i constant to its redundant value w_i.

In this context, the values $w_1,\ldots,w_m$ lie within the *redundancy range* of $W_1,\ldots,W_m$ for the route $(X,Y_1,\ldots,Y_n,Z)$ with actual values $Y_1=y_1, \ldots, Y_n=y_n$, $Z=z$, if the following holds: If for each W_i its equation E_{Wi} is replaced with the new equation $E'_{Wi} := W_i=w_i$ that sets W_i constant to this value w_i, this does not alter the actual values $Y_1=y_1, \ldots, Y_n=y_n$, $Z=z$ of the variables in the route in question.
Notably, the variables $W_1,\ldots,W_m$ in no way have to be intermediate between X and Z, which is a remarkable deviation from my previous reconstructed definition (H AR). In this definition (H WAR), the redundancy criterion only requires that variables $W_1,\ldots,W_m$ are disjoint from the route $(X,Y_1,\ldots,Y_n,Z)$ and that the setting of the values $w_1,\ldots,w_n$ does not alter the values of this selected route for the actual solution $R(V)$ in the model.
My reconstruction of Hitchcock's definition of contributory causation then incorporates this notion of a weakly active route in a completely analogous way as my reconstruction of his definition of actual causation incorporates the notion of an active route.[62]

[62] Cf. [cH 01], p. 289 for Hitchcock's original version of this definition.

Definition (Contributory Causation, Hitchcock) **(H CC)**

Let c and e be distinct occurring events in a situation S, and let $M := \langle V, E \rangle$ be a causal model formalizing S, such that the following holds: There are variables X and Z in V, such that the values x of X and z of Z in the actual solution $R(V)$ of $\langle V, E \rangle$ represent the actually occurring events c and e in S and such that their non-actual values x' and z' represent alterations of c and e respectively.
Then c is a *contributory cause* of e *according to the model* M if and only if in the actual solution $R(V)$ of M there is a weakly active route from X to Z.

The basic idea of Hitchcock's account of actual causation with my reconstructed definitions (H AR) and (HAC) was that c is a cause of e in a situation S if and only if there is a causal process linking these events c and e. A causal process was a concatenation of the mechanisms that were at work in the given situation S and that were linking events c and e. Concretely, Hitchcock formalized a causal process between c and e by an active route in a causal model M between the variables X and Z whose actual values x and z formalized these events c and e. In essence, an active route formally filtered out a chain of mechanisms linking c and e, so that an alteration of event c would propagate along this chain to change the event e. Formally, this isolation of a chain of mechanisms linking c and e was achieved by freezing the other variables Y in the model M that are intermediate between X and Z but do not belong to the route in question. In this way, the mechanisms in situation S not belonging to the causal chain in question were screened off, so that a change in event c could propagate only along the chain to event e.

In this account of contributory causation now with my reconstructed definitions (H WAR) and (H CC), Hitchcock deviates from this rationale of equating a causal process with an active route in a model. Now the formalization of a causal process between a contributory cause c and its effect e is weakly active route in a causal model M between the variables X and Z whose actual values x and z formalize these events c and e. Still, a chain of mechanisms linking c and e in situation S corresponds to this route. However, in the criterion for this route to be weakly active this chain of mechanisms is no longer isolated, so that a change in event c can propagate along other mechanisms outside the chain to event e.

For, in definition (H WAR), the other variables Y in the model M that are intermediate between X and Z but do not belong to the route in question no longer all have to be frozen. Instead, some of these variables Y, denoted by W_i, may be frozen at certain arbitrary values w_i, that may in particular be non-actual, with the only proviso that this freezing does not alter the actual values of the variables along the route.[63] However, there may be others of these variables Y that are intermediate between X and Z but do not belong to the route in question that are not frozen and that retain their original equations. As a result, some of the mechanisms in situation S not belonging to the causal chain corresponding to the route are not screened off, so that a change in event c can also propagate along other mechanisms outside the chain to event e.

I will discuss this difference between my reconstructions of definitions (H AR) and (H WAR) more deeply in the last chapter of this thesis, where I will apply them in the analysis of certain paradigmatic examples. For now, I settle for the remark that (H WAR) is a generalization of (H AR) designed to deal with cases of a special structure by relaxing the rationale

[63] In order to prevent any misunderstandings: Of course, as already mentioned, the variables W_i in the contingency set in definition (H WAR) that are frozen at arbitrary values do not have to be intermediate between X and Z, they just have to be disjoint from the featured route between X and Z.

of isolating causal chains. As a trivial corollary, note that in this way actual causation defined by (H AC) is a special case of contributory causation defined by (H CC).

Let me finally highlight the idea behind this freezing of variables W_i being disjoint from the featured route between X and Z at certain arbitrary values w_i, as long as this freezing does not alter the actual values of the variables along this route: It is exactly this freezing that is intended to screen off the influence of another contributory cause d in the situation S from the effect e. For, observe that for another cause d to symmetrically over-determine the effect e together with cause c, this event d has to work in parallel to c. This means that in the model M formalizing the situation S this additional cause d has to be formalized by a variable W_i that is disjoint from the route between X and Z and not intermediate between X and Z. Freezing W_i at a non-actual value w_i alters the actual event d and graphically removes its influence from the situation. The redundancy of value w_i ensures that in the actual situation the influence of cause c along the featured route to e is not disturbed.

In the following, I will illustrate the application of my above reconstruction of Hitchcock's two definitions (H WAR) and (H CC) with Lewis's classic example 'Firing Squad'. But before this, let me for reasons of fairness state Hitchcock's original formulation of a weakly active route. My reconstruction given above is a slight deviation from it, because I regard Hitchcock's original formulation as misleading. In his original formulation of a weakly active route, Hitchcock suppressed the dependence on a solution of a causal model. This dependence is though vital for a functioning of the definition. However, the reader my compare Hitchcock's original formulation with mine and form a judgment for him- or herself. Hitchcock's original is the following:[64]

[64] Cf. again [cH 01], p. 289 for Hitchcock's original discussion.

Definition (Weakly Active Route, Hitchcock's Misleading Original) **(HMO WAR)**

The route $(X,Y_1,\ldots,Y_n,Z)$ is weakly active relative to $\langle V,E\rangle$, if and only if there exists a set (possibly empty) of variables $\{W_1,\ldots,W_m\}$ in $V \setminus \{X,Y_1,\ldots,Y_n,Z\}$, and values $w_1,\ldots,w_m$ that lie within the redundancy range of $\{W_1,\ldots,W_m\}$ for this route and its actual values, such that Z depends counterfactually upon X within the new system of equations E' constructed from E as follows: for each W_i, replace the equation for W_i with the new equation that sets W_i equal to w_i.

Regarding my reconstruction of definition (H CC), Hitchcock did not give an explicit original formulation, since he only shortly discussed the problem of symmetric over-determination. Instead, Hitchcock just mentioned that his originally formulation of (H AC) has to be carried over analogously, just replacing the notion of an active route with the one of a weakly active route.[65] In the same spirit as in the case of Hitchcock's account of actual causation, I take Hitchcock's original formulations as practical abbreviations of my elaborate reconstructions (H WAR) and (H CC).

Let me now finally come to the discussion of the example 'Firing Squad', which has the following cover story:[66]

Firing Squad: 'A court orders the execution of a prisoner. On the signal of their captain, a two man firing squad with both riflemen being accurate, alert and law-abiding, executes the order. The prisoner dies.'[67]

---

[65] Cf. again [cH 01], pp. 289-290 for Hitchcock's original discussion.
[66] For the original Firing Squad example, cf. section *E* of [dL 86].
[67] The cover story given here is reformulated by me. Pearl himself gives the example with an unknown court order, so that the outcome of the situation is also unknown. Moreover,

For the analysis of this example by my reconstruction of Hitchcock's account with the definitions (H WAR) and (H CC), I suggest the following causal model: There are five bi-valued variables, *U*, *C*, *A*, *B*, and *D*. *U* is the only exogenous variable and has the following meta-assignment of events to its values:

$$U = \begin{cases} 0, & \text{if the court does not order the execution of the prisoner} \\ 1, & \text{if the court does order the execution of the prisoner} \end{cases}$$

Variables *C*, *A*, *B* and *D* are endogenous representing the following event-alterations:

$$C = \begin{cases} 0, & \text{if the captain does not give a signal} \\ 1, & \text{if the captain does give a signal} \end{cases}$$

$$A = \begin{cases} 0, & \text{if rifleman A does not shoot} \\ 1, & \text{if rifleman A does shoot} \end{cases}$$

$$B = \begin{cases} 0, & \text{if rifleman B does not shoot} \\ 1, & \text{if rifleman B does shoot} \end{cases}$$

$$D = \begin{cases} 0, & \text{if the prisoner does not die} \\ 1, & \text{if the prisoner does die} \end{cases}$$

With this, the resulting structural equations for the variables *C*, *A*, *B* and *D* are the following:

$$C = \begin{cases} 0, & \text{if U} = 0 \\ 1, & \text{if U} = 1 \end{cases}, \quad A = \begin{cases} 0, & \text{if C} = 0 \\ 1, & \text{if C} = 1 \end{cases}, \quad B = \begin{cases} 0, & \text{if C} = 0 \\ 1, & \text{if C} = 1 \end{cases}$$

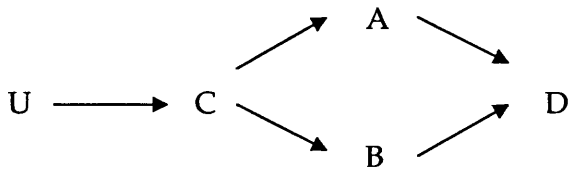
$$\text{and} \quad D = \begin{cases} 1, & \text{if A} = 1 \text{ or B} = 1 \\ 0, & \text{otherwise} \end{cases};$$

or in short form: $C = U$, $B = C$, $A = C$ and $D = \max\{A,B\}$.

---

he explicitly excludes other possible mechanisms that could affect whether the prisoner dies or not.

The graphical representation of 'Firing Squad' finally is:



The actual situation described in the cover story is then modeled by the following value distribution: The exogenous variable U receives the value $U = 1$, yielding the actual solution $C = 1$, hence $A = 1$ and $B = 1$, and in turn $D = 1$.

Suppose we want to analyze whether $B = 1$ is a contributory cause of $D = 1$ in the given situation. I maintain that we have an unequivocal causal intuition in this case. The shooting of rifleman B causally contributes to the death of the victim, if both riflemen A and B shoot. However, we would not say that the shooting of rifleman B actually causes the death of the victim, since rifleman A would still shoot, if rifleman B did not, and the victim would still die.

In this way, I argue, we see a clear difference between actual and contributory causation in his case. Furthermore, the question whether $A = 1$ is a contributory cause of $D = 1$ in the given situation is analogously judged by our pre-theoretic causal intuition in my view: We take the shooting of rifleman A as causally contributing to the death of the victim, but not see it as actually causing it.

Verify first that $B = 1$ is not an actual cause of $D = 1$ under Hitchcock's analysis with definitions (H AR) and (H AC), since the route (B, D) is not active according to (H AR): Changing the value of B to $B = 0$ still brings about $D = 1$. For, the equation E_A of D's other parent A remains unaffected, since A is no descendant of B and hence may not be frozen at its actual value.

However, $B = 1$ is very well a contributory cause of $D = 1$ under the analysis with definitions (H WAR) and (H CC). For, observe that the route (B, D) is weakly active according to (H WAR) in the actual solution of our model given above: Obviously, variable A is disjoint from the route from B to D. And the non-actual value 0 lies in the redundancy range for A and the route (B, D); for, altering the value of A to $A = 0$ does neither affect the value of B nor of D in the actual situation. However, in the modified set of equations E', in which A is set to $A = 0$, an intervention on variable B changing its value to $B = 0$ brings about $D = 0$, i.e. a change in the value of D. With this, route (B, D) is weakly active and according to definition (H CC), $B = 1$ is a contributory cause of $D = 1$.
With an exactly analogous argumentation one can also show that the route (A, D) is weakly active according to (H WAR) in the actual solution of our model given above, but not active according to (H AR). In this way, $A = 1$ is also a contributory cause of $V = 1$, but no actual cause, in Hitchcock's analysis.

Consider now, how Lewis's classic counterfactual account would analyze this example 'Firing Squad': Let b_1 denote the event that rifleman B shoots, corresponding to $B = 1$ in our causal model above. Analogously, let a_1 and v_1 denote the events that rifleman A shoots and that the victim dies respectively, corresponding to $A = 1$ and $V = 1$ respectively in Hitchcock's causal model.
Discuss first whether b_1 actually causes v_1 according to Lewis's definition (L AC). There are no intermediate events between b_1 and v_1, so that there is only the trivial chain consisting of events b_1 and v_1 themselves that connects b_1 and v_1. In this way, b_1 is an actual cause of v_1 according to (L AC) if and only if v_1 is causally dependent on b_1 according to (L CD). And according to our definition of causal dependence, v_1 was causally dependent on b_1 if and only if the following counterfactual held: If b_1 had

not occurred, then v_1 would not have occurred. We detect that if b_1 had not occurred, then v_1 would still have occurred, since a_1 still occurred.

Note that the non-backtracking requirement for definition (L CD) decisively enters here. In evaluating the counterfactual 'If b_1 had not occurred, then ...' we keep the event c_1 fixed that denotes the order of the captain to fire, corresponding to $C = 1$ in our causal model. If the change of event b_1 were allowed to track back to this event c_1, on which b_1 causally depends, this backtracking would alter the result. Then with the captain not ordering to shoot, rifleman A would not shoot as well, so that the victim would survive. However, with prohibiting backtracking, we have the result that b_1 is not an actual cause of v_1 according to (L AC).

With an analogous argumentation one can show that a_1 is not an actual cause of v_1 according to (L AC) as well.

In this way, the verdicts that Lewis's account delivers stand in a clear contradiction to what I take to be our causal intuitions about this example. Lewis tried to avoid this problem by simply claiming that he did not have any clear intuitions about such cases, i.e. cases in which two or more events symmetrically over-determine an effect.[68] However, I view this more as a desperate move than as a proper solution of the problem. In my opinion, we have clear intuitions about this example, and see a clear difference between actual and contributory causation. Not being able to account for these intuitions cannot be satisfactory for Lewis.

As a result, the last of my featured three advantages of Hitchcock's account in comparison to Lewis's account becomes apparent: Hitchcock's account analyzes a case symmetric over-determination in accord with our causal intuition, whereas Lewis's account fails to do this.

[68] Cf. again section *E* of [dL 86].

WOODWARD'S ACCOUNT OF INCLUSIVE TOKEN CAUSATION BEING EQUIVALENT TO HITCHCOCK'S ACCOUNT OF CONTRIBUTORY CAUSATION

In this section I will give a brief synopsis of the basic facts of Woodward's account of token causation that is featured in Woodward's influential book *Making Things Happen*.[69] In a certain way, this presentation will do injustice to Woodward, because he does not really intend to give an account of his own for dealing with token causation. Only a comparatively small part of the book deals with modeling token causation. Instead, *Making Things Happen* gives an overview of an interventionist theory of causation and explanation. It touches on topics like the notion of an intervention and its justification, the notion of invariance and the way in which it offers a new understanding of the concept of a law in causal explanations, and the issue of how causal relationships can be exploited for purposes of manipulation and control. In my view, the book more than succeeds here, and it does so in a very illuminating way.

Concerning the issue of modeling token causation, Woodward's intention is merely to give a sketch of certain characteristics that a successful account of token causation has to fulfill in his view. For many details of this sketch Woodward explicitly refers to Hitchcock's article 'The Intransitivity of Causation Revealed in Equations and Graphs', with which the other sections of this chapter have dealt.[70]

My presentation and discussion of Woodward's sketch that I will offer in this section will be very short compared to the way in which I have dealt

[69] Cf. [jW 03].
[70] Cf. for example [jW 03], p. 83.

with Hitchcock. I see my justification for this mode of presentation in the following fact: The relevant definitions in Woodward's sketch of token causation are all taken over from Hitchcock. In particular Woodward's final definition of token causation that I will denote (W TC) is equivalent to Hitchcock's definition of contributory causation (H CC). In this way, most of the remarks that I have made about Hitchcock's account in the other sections of this chapter equally apply to Woodward's account. My reason for focusing on Hitchcock's account of token causation in this chapter is mainly that Woodward explicitly refers to it, that it is more detailed, and finally that it has been in strong need of a conceptual clarification.

In strong contrast to Hitchcock, Woodward does not distinguish between actual and contributory causation in his account. For Woodward, there is only one inclusive notion of singular event causation that also applies to cases of symmetric over-determination. However, Woodward's definition of token causation (W TC) is equivalent to Hitchcock's definition of contributory causation (H CC) employing the notion of a weakly active route defined by (H WAR). The difference from Hitchcock is that Woodward does not limit the extension of his definition (W TC) to cases of symmetric over-determination, as Hitchcock does, but that Woodward takes his definition to cover all cases of singular event causation. However, for reasons of fairness, I have to mention that the term 'token causation' is not Woodward's original terminology. Instead, Woodward calls the concept with which this section will deal 'actual causation'. I have chosen this way of talking because I consider it to be more suggestive and also to avoid confusion.

Since I consider my remarks about the relevant concepts in Hitchcock's account of token causation to translate into results about the analogous concepts in Woodward's account, I mainly aim to show that the concepts Woodward employs are equivalent to Hitchcock's.

I will confine myself to a short demonstration that Woodward's definition of token causation (W TC) is equivalent to my reconstruction (H CC) of Hitchcock's definition of contributory causation. For this aim I will also mention a preliminary version of this definition that Woodward gives for heuristic reasons, and I will argue that this preliminary version is equivalent to my reconstruction (H AC) of Hitchcock's definition of actual causation.

I will save the reader a lengthy argumentation that the relevant concepts in the modeling procedure, i.e. the definitions of a causal model, of counterfactual dependence, of exogeneity, endogeneity and structural equations, and of parenthood and childhood of variables, which I have discussed in the first section of this chapter, all translate into Woodward's modeling framework for analyzing token causation. Instead, I refer the reader to Woodward's original discussion in order to verify that my definitions (H CM), (H CD), and (H EESE) all have equivalent analogues in Woodward's account, even if only implicitly, and that also my reconstruction of the modeling procedure for analyzing token causation by steps M1) and M2) is compatible with Woodward's account. Admittedly, Woodward does not explicitly deal with the generation of causal models for analyzing token causation, and moreover he avoids precise formal definitions of most of these concepts. This is due to the fact that the modeling of token causation is only of minor importance for him. Nevertheless, I maintain that his understanding of the respective notions can be extracted from his discussion more or less straightforwardly.[71]

---

[71] In this way, cf. [jW 03], pp. 42-45 and pp. 327-336 for Woodward's notion of a causal model for the analysis of token-level causal dependencies, also called a 'system of structural equations' by him. For his definition of counterfactual dependence, also called 'direct causation' by him, cf. [jW 03], p. 55. For Woodward's understanding of structural equations, cf. [jW 03], pp. 42-48, pp. 52-53, and again pp. 327-336. For my assertion that Woodward's sketchy modeling framework for analyzing token causation is compatible with steps M1) and M2), cf, again [jW 03], pp. 42-45 and also pp. 74-76. Finally, for Woodward's discussion of the notion of an intervention, cf. [jW 03], pp. 94-114.

Let me start my discussion with some broad considerations about unified token causation. Certainly such an understanding of token causation as a unified notion comprising both actual and contributory causation has appeal. After all, Hitchcock's basic rationale of a causal process linking token cause and token effect is the same in both cases.

The difference between actual and contributory causation for Hitchcock is located in the kind of circumstances in which a causal process operates. For actual causation this process has to operate in the very circumstances prevailing in the situation to be modeled: Hitchcock freezes the variables not immediately involved in the process at their actual values. For contributory causation, he allows the process to operate in non-actual circumstances that correspond to freezing variables not immediately involved in the process at arbitrary values, with the only constraint being that this does not disturb the actual value of the effect variable.

Also the formal definitions (H AC) and (H CC) of actual causation and contributory causation are related: We saw this in the previous section of this chapter, when we noticed that an active route defined by (H AR) is a special case of a weakly active route defined by (H WAR), so that for Hitchcock actual causation is a special case of contributory causation.

The starting point of Woodward's discussion of modeling token causation is the construction of the following preliminary account.[72]

Definition (Token Causation Preliminary, Woodward) **(W TCP)**

$X=x$ is a *preliminary token cause* of $Z=z$ if and only if the following two conditions are both satisfied:

(i) The actual value of X is x and the actual value of Z is z.

[72] Cf. [jW 03], p. 77. My description here is very close to Woodward's original in order to give the reader a better impression of the comparatively low level of formalization in Woodward's account.

(ii) There is at least one route r from X to Z for which an intervention on X will change the value of Z, given that other *direct causes* Y_i of Z that are not on this route have been fixed at their actual values.
(It is assumed that all direct causes of Z that are not on any route from X to Z remain at their actual values under the intervention on X.)

In this context, Woodward takes a *direct cause* to be defined as follows:[73]

Definition (Direct Causation, Woodward) **(W DC)**

A necessary and sufficient condition for X to be a *direct cause* of Z with respect to some variable set V is that there be a possible intervention on X that will change Z when all other variables in V besides X and Z are held fixed at some value by interventions.

I maintain that this condition (W DC) is in essence a less formal expression of my reconstruction of Hitchcock's definition of counterfactual dependence of variables in definition (H CD). In order to verify this, suppose that $V = \{X, Y_1, \ldots, Y_n, Z\}$ is the set of variables selected to be relevant for the analysis of causal dependencies in a given situation S. In my view, the condition in (W DC), i.e. that there is an intervention on X that will change Z when all other variables in V besides X and Z are held fixed at some value by interventions, is equivalent to the following condition: There are values $x \neq x'$ of X and $z \neq z'$ of Z, and values y_i for the other variables Y_i, such that

$X = x \;\&\; Y_1 = y_1 \;\&\; \ldots \;\&\; Y_n = y_n \;\boxed{\rightarrow}\; Z = z$ and

$X = x' \;\&\; Y_1 = y_1 \;\&\; \ldots \;\&\; Y_n = y_n \;\boxed{\rightarrow}\; Z = z'$.

[73] Cf. [jW 03], pp. 42 and 55.

And this is the defining condition of counterfactual dependence according to my reconstruction of Hitchcock's definition (H CD). The only difference between these two formulations (H CD) and (W DC) is their ontology: Woodward talks of interventions in (W DC), whereas my reconstruction of Hitchcock in (H CD) employs counterfactuals. Apart from this, (H CD) and (W DC) express the same condition, i.e. Z is counterfactually dependent on X, given V, according to (H CD) if and only if X is a direct cause of Z, given V, according to (W DC). This difference in ontology concerns the procedure of constructing a causal model M suitable for describing the situation S, with which Woodward does not deal. In this way, I propose to leave this issue aside and to settle for the equivalence of (H CD) and (W DC) in principle.[74]

With this, I come to Woodward's preliminary definition of token causation for heuristic reasons (W TCP). I maintain that (W TCP) expresses in less formal terminology more or less exactly the condition that is required in my reconstruction of Hitchcock's definition of actual causation (H AC). In order to scrutinize this, let the prerequisites of (H AC) be met, i.e. let c and e be distinct occurring events in a situation S, and let $M := \langle V, E \rangle$ be a causal model formalizing S, such that the following holds: There are variables X and Z in V, such that the values x of X and z of Z in the actual solution $R(V)$ of $\langle V, E \rangle$ represent the actually occurring events c and e in S and such that their non-actual values x' and z' represent alterations of c and e respectively. On the foil of this prerequisite, the defining condition in (H AC), namely that c is an actual cause of e according to the model M if and only if in the actual solution $R(V)$ of M there is an active route from X to Z is equivalent to clauses (i) and (ii) in (W TCP). First, observe that with the prerequisite, clause (i) of (W TCP) is already fulfilled. Second,

[74] Nevertheless, I still maintain that a basic set of super-exhaustive counterfactuals is a straightforward way of describing the complete counterfactual information about a situation S that a causal model M has to incorporate,

according to definitions (H AR) and (H CDSCM) a route r from X to Z is active in a solution $R(V)$ of the causal model $M = \langle V, E \rangle$ if and only if an intervention on X will change the value of Z, when all intermediate variables W_j between X and Z that are not on route r have been fixed at their actual values and the structural equations of all other variables in V remain the same. Third, observe that the value of Z is determined only by the parent variables of Z and not by more remote ancestor variables, and that the parents of Z are exactly its direct causes, as can be seen in definition (W DC). Fourth, for this reason, fixing all intermediate variables W_j between X and Z that are not on route r at their actual values, as demanded by (H AR) and (H CDSCM) in this context, brings about exactly the same result for variable Z as fixing only the intermediate variables U_j between X and Z that are not on route r and direct causes of Z, as demanded by clause (ii) of (W TCP). Analogously fifth, the remaining of the direct causes Y_i of Z that are not on any route from X to Z at their actual values, as demanded by clause (ii) of (W TCP) in this context, brings about exactly the same result for variable Z as keeping the structural equations of all variables in V that are not intermediate between X and Z, as demanded by (H AR) and (H CDSCM). All in all, on the foil of the mentioned prerequisite, the condition in (H AC) is equivalent to clauses (i) and (ii) of (W TCP). In this way, Woodward's preliminary definition (W TCP) is a less formal expression of exactly the underlying rationale that I attributed to Hitchcock, namely to identify a causal process in a situation S with a corresponding route in an appropriate model M.

Admittedly, Woodward's preliminary definition (W TCP) does not directly mention models M and hence seems to canonically account for a notion of causation that is independent of our description. However, with the same argumentation as in the fourth section of this chapter applying to Hitchcock's account, I consider such an undertaking, if it should be intended by Woodward, to be misguided. I still maintain that relativizing

token causation to a given model $\boldsymbol{M}$ allows us to analyze the causal relationships between exactly the kinds of events in which we are interested in a situation *S*.

A side observation that can be made concerning Woodward's definitions (W TCP) and (W DC) is that they seem to use interventions in two different ways. In definition (W TCP), interventions operate on a set of structural equations, whereas in definition (W DC), interventions operate in a set of variables, before structural equations have been set up. Instead, it is these interventions that allow setting up structural equations. Woodward gives an illuminating formal account of interventions, in which he describes what conditions an intervention has to meet so that it can formalize an ideal experimental manipulation on the reference object of a variable X. However, a discussion of this account would lead me too far astray here. Thus, I would like to refer the reader to Woodward's original remarks.[75]

As mentioned, for Woodward, the above preliminary definition (W TCP) is just a heuristic step on the way to a more adequate definition of token causation. Woodward correctly observes that (W TCP) cannot account for cases of symmetric over-determination,[76] an observation that Hitchcock also made about his account of actual causation. However, Woodward draws a different conclusion than Hitchcock. Instead of devising a separate account for contributory causation applying to exactly these cases, Woodward wants to extend his preliminary definition of token causation to uniformly cover all cases of singular event causation. In this way, no matter whether an effect *e* is over-determined by two or more causes *a* and *c* that work in parallel or whether there is no such over-determination, such that for example there is just one cause *d* for *e*, Woodward champions the intuition that the over-determining causes *a*

[75] Cf. [jW 03], pp. 98-107.
[76] Cf. [jW 03], p. 82.

and c are token causes of e in the same sense as d is.[77] I call the understanding of causation that is expressed by this intuition *Woodward's inclusive understanding of token causation*. Woodward proposes the following semi-formal definition of this inclusive notion.[78]

Definition (Token Causation, Woodward) **(W TC)**

$X = x$ is a *token cause* of $Z = z$ if and only if the following two conditions are both satisfied:

(TC1) The actual value of X is x and the actual value of Z is z.

(TC2) For each route r from X to Z, fix by interventions all direct causes Y_i of Z that do not lie along r at some combination of values within their redundancy range. Then determine whether, for each route from X to Z and for each possible combination of values for the direct causes Y_i of Z that are not on this route and that are in the redundancy range of Y_i, whether there is an intervention on X that will change the value of Z. Clause (TC2) is satisfied if and only if the answer to this question is 'yes' for at least one route and possible combination of values within the redundancy range of the Y_i.

As in the case of the preliminary definition of token causation (W TCP), also this definition (W TC) is borrowed from Hitchcock. The difference is that this time Woodward explicitly refers to Hitchcock's original. He does this on the occasion of taking over Hitchcock's definition of a redundancy range.[79]

---

[77] Cf. [jW 03], pp. 82-83.
[78] Cf. [jW 03], pp. 83-84. Again, my description almost literally Woodward's original.
[79] Cf. [jW 03], p. 83.

Let me close with my main claim in this section, namely that this definition (W TC) is a less formal expression of my reconstruction of Hitchcock's definition of contributory causation (H CC) employing a weakly active route (H WAR). My argument proceeds in exactly the same way as with definition (W TCP). Again, let the prerequisites of (H CC) be met, i.e. let c and e be distinct occurring events in a situation S, and let $M := \langle V, E \rangle$ be a causal model formalizing S, such that the following holds: There are variables X and Z in V, such that the values x of X and z of Z in the actual solution $R(V)$ of $\langle V, E \rangle$ represent the actually occurring events c and e in S and such that their non-actual values x' and z' represent alterations of c and e respectively. On the foil of this prerequisite, the defining condition in (H CC), namely that c is a contributory cause of e according to the model M if and only if in the actual solution $R(V)$ of M there is a weakly active route from X to Z is equivalent to clauses (i) and (ii) in (W TC). Again, with this prerequisite, clause (i) of (W TC) is already fulfilled. And according to definitions (H WAR) and (H CDSCM), a route r from X to Z is weakly active in a solution $R(V)$ of the causal model $M = \langle V, E \rangle$ if and only if the following condition is fulfilled: There is a set $\{W_1,..,W_n\}$, possibly empty, of redundant variables W_i that do not lie along r and a combination of values $(w_1,..,w_n)$ for these W_i within their redundancy range, such that an intervention on X will change the value of Z if these W_i are frozen at these values w_i. Again, freezing the whole set of redundant variables W_i at their redundant values brings about the same result for variable Z as merely freezing the subset of $\{W_1,..,W_n\}$ that consists of exactly the direct causes of Z. For, again, the value of Z is determined only by the parent variables of Z and not by more remote ancestor variables. All in all, on the foil of the mentioned prerequisite, the condition in (H CC) employing (H WAR) is equivalent to clauses (i) and (ii) of (W TC).

Again, this definition (W TC) of Woodward's does not really mention models. However, with the same argumentation as before I still maintain that relativizing token causation to a model M offers a pragmatically more successful approach to analyzing a wide range of situations S with their constituting events for causal relationships.

CHAPTER 2

PEARL'S ACCOUNT OF MODELING TOKEN CAUSATION

This chapter is devoted to a thorough discussion of Pearl's account of modeling token causation. Coming from an engineering background and not being a philosopher by training, Judea Pearl started his study of causation from an empiricist tradition. Pearl mainly conducted research in probability theory and artificial intelligence and had the conviction that causal relationships were just useful abbreviations of probabilistic patterns. Pearl's main work in this field is his book *Probabilistic Reasoning in Intelligent Systems,*[80] in which he maintained that probabilistic relationships are the foundation for our understanding of the world and that causation is just a secondary and more graphical concept.[81] Since then, Pearl's view of causation has changed dramatically. In his book *Causality – Models, Reasoning, and Inference,*[82] which is commonly regarded as a milestone in the study of causation, Pearl regards the probabilistic relationships as mere surface phenomena and takes the causal relationships as the fundamental building blocks both of physical reality and of our understanding of this reality.[83] In my view, Pearl's work on causation culminated in this later book *Causality,* and without a doubt it has been more influential in the philosophical community than any other of his publications. What remains from Pearl's engineering background is a rather high level of formal apparatus in his account of causation.

In this chapter, I focus on the account of modeling token causation that Pearl gives in *Causality.* In this book, Pearl gives accounts of many aspects of causal discourse. Pearl deals with causation on both the type and the token level, discusses the problem how to derive causal models from raw data and gives examples of plenty of applications in various fields of science.

---

[80] Cf. [jP 88].
[81] Cf. the introduction of [jP 88].
[82] Cf. [jP 00].
[83] Cf. the introduction of [jP 00].

Regarding causation on the token level, Pearl discusses two concepts: The first is *in fact token causation*, i.e. the relation that holds between two singular events c and e, when c in fact causes e.[84] The second is the *probability of necessity, or sufficiency respectively, of token causation*, i.e. the probability that an event c was a necessary cause, or sufficient cause respectively, of another event e.[85]

In this chapter, as in the whole thesis, I only deal with *in fact token causation*, which I have been denoting, and will continue to denote, shortly as *token causation*. In my view in fact token causation is the more paradigmatic concept of the above, for more than one reason. First, we have the strongest pre-theoretic causal intuitions when we are asked to judge whether in a particular situation S a certain event c in fact causes another event e. Second, from an explanatory point of view, a concept of in fact token causation is the purest and simplest to convey, since it does not involve other problematic notions like necessity and sufficiency. And finally, from a pragmatic point of view, the formal apparatus required to analyze in fact token causation is the most developed so far.

Furthermore, the focus in this chapter is, as in the whole thesis, on the second problem of causal modeling, i.e. the definition of token causation relativized to a given causal model. The first problem of causal modeling, i.e. the derivation of an appropriate causal model from a given situation, is only dealt with marginally.

[84] Cf. [jP 00], chapter 10.
[85] Cf. [jP 00], chapter 9.

The first section deals with the basic building block of Pearl's account of token causation, his definition of a causal model. In essence, his causal model M is devised to analyze a given situation S that contains certain token events *c*, *d*, *e*, etc., in whose causal relationships we are interested. For this purpose, Pearl's causal model M is comprised of two sets of variables, exogenous variables $\boldsymbol{U}$ and endogenous variables V. The exogenous variables $\boldsymbol{U}$ formalize admissible alterations of background events in the situation S, whose development is either unknown to us, or at least not represented in the causal model M. Analogously, the endogenous variables V formalize admissible alterations of the events in situation S, whose development represented in the causal model M. A set of functions F in the causal model M is devised to determine the values of exactly these endogenous variables V from the values of the exogenous ones. The underlying rationale is that these functions F formalize deterministic mechanisms, or laws, that are at work in the situation S. Since Pearl's way to define his deterministic causal models is, apart from terminological differences, in essence the same as for Hitchcock, I refrain from an elaborate description of Pearl's definition and instead try to explain and motivate it by working out the differences to Hitchcock's definition of a causal model. Remarkably, and in contrast to Hitchcock's account, Pearl's definition of a causal model is able to serve not only in an analysis of token causation, but can also be employed to analyze causal claims on the type level. For this reason, I also briefly describe the probabilistic extension of Pearl's deterministic causal model in this section. The next sections of this chapter are then devoted to the peculiarities of Pearl's account of token causation, and in their discussion and motivation I will no longer draw any comparisons to Hitchcock. Instead, the next chapter is devoted to a thorough comparison of their accounts.

In the second section, I discuss the main concept that underlies Pearl's account of token causation, the concept of sustenance. Pearl's basic idea is that for an event c to qualify as a cause for another event e in an arbitrary situation S, event c has to be able to sustain event e against certain contingencies. As will turn out in this discussion, these contingencies are structural in nature, which means that they correspond to surgical interventions in a causal model $\boldsymbol{M}$ that is employed for formalizing the situation S that is to be causally analyzed. Because of this decisive entering of interventions in his conception of sustenance and also because of the counterfactual information encoded in the functions in Pearl's models, his account of token causation can rightfully be classified as both counterfactual and interventionist.

The third section clarifies in precisely what form a cause c has to sustain its effect e in a given situation S with a certain constellation of background conditions. I maintain that this can be most easily understood with the aid of the leading idea that c sustains its effect e via a causal process that links cause c and effect e. I motivate Pearl's definitions of a causal and a natural beam as the decisive step in reaching a formalization of this notion of a sustaining causal process. Causal or natural beams $\boldsymbol{M_u}$ are projections of a causal model $\boldsymbol{M}$ on its actual state $\boldsymbol{U = u}$ that formalizes the constellation of background conditions in the given situation S that is to be analysed for causal dependencies. I will show in this section that a causal or a natural beam $\boldsymbol{M_u}$ describe under which conditions a set of parent variables S_i locally sustains the actual value v_i of their child variable V_i, and this for every endogenous variable V_i in the causal model $\boldsymbol{M}$ in the state $\boldsymbol{u}$. Pearl's notion of sustenance though also gives guidelines on what it means for the value x of an arbitrary variable X to sustain a value y of another arbitrary variable Y. These guidelines are put into concrete terms by Pearl's definitions of actual and contributory causation, with which I deal in the fourth section.

The fourth section is devoted to a thorough discussion of Pearl's definitions of actual and contributory causation, which are the two kinds of token causation he distinguishes. Pearl's definition of actual causation decisively employs the notion of a natural beam, whereas his definition of contributory causation does the same with the notion of a non-natural causal beam. In essence, x is called an actual cause, or contributory cause respectively, of y in a causal model M in the state u, if and only if x sustains y against certain structural contingencies in a natural beam, or in a non-natural causal beam. Pearl intends his definition of contributory causation to formalize a weakened notion of actual causation that exclusively applies to cases of symmetric over-determination. I argue that such a distinction between two mutually disjoint kinds of token causation is not necessary, and I offer a trivial unified definition of token causation that contains actual causation as a special case. Furthermore, I argue that Pearl's definitions of actual and contributory causation only give a pragmatic account of what token causation relatvized to a given model amounts to. Finally, I maintain that a causal process can be understood to be the concatenation of the projection functions f_i^u in a causal or natural beam M_u of exactly the variables V_i that are intermediate between the cause variable X and the effect variable Y.

In the fifth section, I illustrate how a natural beam is constructed. I focus on the discussion of natural beams since they are simpler and in a certain way more paradigmatic than non-natural causal beams. A remarkable observation that I make in this illustration is that treatment of exogenous variables in constructing local sustaining sets is not unequivocally covered by Pearl's definition of a natural beam.

In the sixth section, I point out striking differences between Pearl's formalization of a causal process with the aid of natural and causal beams on the one hand and Hitchcock's formalization of a causal process by a simple causal route on the other. Moreover, I draw attention to the fact that values of exogenous variables can neither be actual nor contributory causes for Pearl, mainly since he excludes exogenous variables from sustaining sets in the construction of both a natural and a causal beam.

In the seventh section, I give a brief synopsis of how Pearl generates his causal models and compare this to the way in which Hitchcock arrives at his causal models for a given example situation. Pearl's approach to generating causal models is mainly geared to analyzing causal relationships on the type level. Pearl's starting point is the definition of conditional probabilistic dependence for a set of random variables $T := \boldsymbol{U} \cup \boldsymbol{V}$ with a joint probability distribution P extending over all admissible, in particular over all non-actual, combinations of values of the variables in $\boldsymbol{U} \cup \boldsymbol{V}$. Pearl utilizes this probability distribution P and this definition of conditional probabilistic independence in his so-called algorithm of inductive causation to determine an equivalence class $[D]$, a so-called pattern, of basic causal structures D that can all generate this probability distribution P. In this way, the result of Pearl's model generating procedure is not a complete causal model M with a set of deterministic functions F, but instead, the result is the equivalence class $[D]$ of basic causal structures D that are in essence just the set of variables $\boldsymbol{U} \cup \boldsymbol{V}$ together with a relation R describing parenthood between these variables.

Admittedly, my synopsis is in no way complete nor is it self-contained. Instead, I would like to refer the reader to Pearl's own discussion of the model generation procedure, which is far more thorough and comprehensive. I mention Pearl's model generation procedure merely for the following reasons:

First for drawing the conclusion that in the application of Pearl's definitions of actual or contributory causation, the causal model M out of which the beams M_u are constructed is strongly underdetermined, because the set of functions F determining the values of the variables in V has to be stipulated. Second, to compare Pearl's approach with the one of Hitchcock.

Apart from this, the only critical remark that I have to make regarding Pearl's account of model generation is that his stipulation of a probability distribution P over a set of random variables $U \cup V$ that formalize possible alterations of *singular* events could be problematic and is definitely in need of motivation.

In a final comparison with Hitchcock's account of model generation, the following striking difference becomes apparent: Pearl has a strictly pragmatic strategy employing probability distributions to extract causal patterns out of them and ending up with a multitude of basic causal structures that can all serve as blueprints for a complete causal model. Hitchcock's strategy is more metaphysical in nature because the complete counterfactual information about a situation is stipulated at the outset of model construction and uniquely and completely determines a causal model on which Hitchcock's definitions of actual or contributory causation are applied.

In this way the advantage of a causal modeling account over classic metaphysical accounts of causation, namely to divide the analysis of causation into two sub-problems – first the generation of a suitable causal model to formalize a given situation and second the definition of causation qualified to such a model – is lost. As a result, I finally maintain that Hitchcock fails in his attempt to simplify Pearl's account of the first problem of causal modeling, namely the generation of a suitable causal model to formalize a given situation S.

PEARL'S DETERMINISTIC CAUSAL MODELS FOR TOKEN-LEVEL CAUSATION AND THEIR PROBABILISTIC EXTENSION

This section deals with the basic building block of Pearl's account of token causation, his definition of a causal model. In essence, his causal model M is devised to analyze a given situation S that contains certain token events c, d, e, etc., in whose causal relationships we are interested. For this purpose, Pearl's causal model M is comprised of two sets of variables, exogenous variables $\boldsymbol{U}$ and endogenous variables V. The exogenous variables $\boldsymbol{U}$ formalize admissible alterations of background events in the situation S, whose development is unknown to us. Analogously, the endogenous variables V formalize admissible alterations of the events in situation S, about whose development we are not ignorant. A set of functions F in the causal model M is devised to determine the values of exactly these endogenous variables V from the values of the exogenous ones. The underlying rationale is that these functions F formalize deterministic mechanisms, or laws, that are at work in the situation S.

Since Pearl's way to define his deterministic causal models is, apart from terminological differences, in essence the same as for Hitchcock, I refrain from an elaborate description of Pearl's definition and instead try to explain and motivate it by working out the differences to Hitchcock's definition of a causal model. Remarkably, and in contrast to Hitchcock's account, Pearl's definition of a causal model is, though, able to serve not only in an analysis of token causation, but can also be employed to analyze causal claims on the type level. For this reason, I also briefly describe the probabilistic extension of Pearl's deterministic causal model in this section.

The next sections of this chapter will be devoted to the peculiarities of Pearl's account of token causation. In their discussion and motivation I will no longer draw any comparisons to Hitchcock. Instead, the next chapter is devoted to a thorough comparison of their accounts. The question finally, how Pearl generates a suitable causal model $\boldsymbol{M}$ to formalize a given situation S will be addressed in the last section of this chapter.

Pearl's formal definition of a causal model is now the following:[86]

Definition (Causal Model, Pearl) **(P CM)**

A causal model is a triple $\boldsymbol{M} = \langle \boldsymbol{U}, \boldsymbol{V}, \boldsymbol{F} \rangle$, where

(i) $\boldsymbol{U}$ is a set of background variables, also called *exogenous*, whose values are determined by factors outside the model;

(ii) $\boldsymbol{V}$ is a set $\{V_1,\ldots,V_n\}$ of variables called *endogenous*, whose values are determined by variables in the model, i.e. by variables in $\boldsymbol{U} \cup \boldsymbol{V}$;

(iii) $\boldsymbol{F}$ is a set of functions $\{f_1,\ldots,f_n\}$, such that each f_i is a mapping from the respective value-domains of $\boldsymbol{U} \cup (\boldsymbol{V} \setminus V_i)$ to that of V_i and such that the entire set $\boldsymbol{F}$ forms a mapping from $\boldsymbol{U}$ to $\boldsymbol{V}$ and the entire set $\boldsymbol{F}$ has a unique solution.

[86] Cf. [jP 00], p. 203.

Symbolically, the set of equations $\boldsymbol{F}$ can be represented by writing

$$v_i = f_i(\boldsymbol{pa}_i, \boldsymbol{u}_i), \quad i = 1,\ldots, n,$$

where $\boldsymbol{pa}_i$ is any realization of the unique minimal set of variables $\boldsymbol{PA}_i$ in $\boldsymbol{V} \setminus V_i$ sufficient for representing f_i. Likewise, $\boldsymbol{U}_i \subseteq \boldsymbol{U}$ stands for the unique minimal set of variables in $\boldsymbol{U}$ sufficient for representing f_i.

One can identify Pearl's notions of exogeneity, endogeneity and parenthood with Hitchcock's respective notions introduced in the last chapter. There is only one important difference, namely that Hitchcock establishes the exogeneity, endogeneity, and the parent relationship for the variables that he employs starting out from a basic set $\boldsymbol{C}$ of underlying counterfactuals, whereas Pearl stipulates the exogeneity, endogeneity, and the parent relationship for his variables directly in his causal models.

Pearl's functions $f_1,\ldots,f_n$ in set $\boldsymbol{F}$ are just a more natural notation for Hitchcock's structural equations. I already remarked in the last chapter that Hitchcock's structural equations are not really equations, but functions that encode the counterfactual relationships between the argument variables and the image variable. The same difference can be found here again, namely that Hitchcock's structural equations are condensed out of the complete counterfactual information contained in this set $\boldsymbol{C}$ of basic counterfactuals, whereas for Pearl the functions $f_1,\ldots,f_n$ in set $\boldsymbol{F}$ are stipulated directly in the causal models. Apart from that, for both Pearl and Hitchcock these functional relationships f_i share the underlying rationale of representing the mechanisms, or laws, that are at working in the situation that is to be modeled and causally analyzed. I will deal with Pearl's methodology of generating his causal models more elaborately in the last section of this chapter. For now, let me settle for some more technical remarks about the above definition (P CM):

First, for an arbitrary sequence W of variables, a *realization* w is a sequence of values of the respective variables in W. Second, a set of variables X is *sufficient* for representing a function $y = f(x, z)$, if $f(x, z)$ is independent of the set of variables $Z := V \setminus X$, i.e. if for all realizations x of X and z, z' of Z we have $f(x, z) = f(x, z')$.[87] Finally, PA_i and U_i are the intersections with the set of endogenous variables, and the set of exogenous variables respectively, of the unique minimal set $T_i := PA_i \cup U_i$ that is sufficient for representing f_i – neither PA_i nor U_i are individually sufficient in this respect; Pearl's description is a bit misleading here.[88]

A very remarkable feature of Pearl's causal models $M = \langle U,V,F \rangle$ defined by (P CM) is that they can be employed to model general as well as singular causation. According to Pearl, this classification into general and singular causation depends on the amount of scenario-specific information that is required to evaluate these causal claims. For Pearl, causal claims are categorized as singular or token-level if they refer to information about special events in a certain scenario, whereas they are classified as general or type-level if they are relative to types of events. In definition (P CM), such a scenario is formalized by model state $U = u$, which describes all relevant details of the situation at hand, and which completely determines the value distribution, a so-called solution, for all endogenous variables V in the model. In particular, this means that Pearl's causal models as defined by (P CM) are clearly deterministic.

---

[87] Again with the same qualification as in the above footnotes, observe that this notion of sufficiency is equivalent to Hitchcock's definition of counterfactual dependence of variables (H CD).

[88] The variables PA_i can be understood to be simply the parents of variable V_i that are endogenous, whereas U_i are the parents of V_i that are exogenous in this understanding. To be accurate, Pearl does not give an explicit definition of parenthood. However, I take it he assumes the following: Variable X is a parent of variable Z iff the values of X occur in the argument of f_Z.

Another motivation for the distinction between token-level and type-level causation is given according to Pearl by the fact they have different sources for their explanatory power: If cause x receives its explanatory power by its tendency to produce y, compared to the weaker tendencies of the alterations of x to produce y, we have a case of general causation. If the explanatory power of cause x is due to x being necessary for triggering a specific chain of events leading to y in the specific situation at hand, we have a case of singular causation.[89]

In cases of general causation, model state $\boldsymbol{U} = \boldsymbol{u}$ is not determined for Pearl, but there is a natural measure for the probability that a certain state prevails. This measure is given by a probability function $P(\boldsymbol{u})$ over the possible model states. Such a probability function $P(\boldsymbol{u})$ then also induces a probability distribution for the set of endogenous variables $\boldsymbol{V}$. Pearl hence extends his definition of a causal model (P CM) by a probability function, in order to deal with cases of general causation. This extension is given by his following definition:[90]

Definition (Probabilistic Causal Model, Pearl) **(P PCM)**

A probabilistic causal model is a pair $\langle \boldsymbol{M}, P(\boldsymbol{u}) \rangle$, where $\boldsymbol{M}$ is a causal model and $P(\boldsymbol{u})$ is a probability function defined over the domain of $\boldsymbol{U}$.

This probability function $\boldsymbol{P}$ then also induces a probability distribution over the whole set $\boldsymbol{V}$ of endogenous variables in $\boldsymbol{M}$, given by

$$P(y) \equiv P(Y = y) = \sum_{\{u \mid Y_x(u) = y\}} P(u).$$

[89] For Pearl's distinction between general and singular causation, cf. [jP 00], p. 222 as well as pp. 253 – 256 and pp. 309 and 310.
[90] Cf. [jP 00], pp. 205 and 206.

I do not intend to pursue this topic of probabilistic causal models and their applications in dealing with cases of general causation further in this thesis. I mentioned this definition of a probabilistic causal model mainly, because it motivates two things: First, that claims of token causation are extreme or marginal cases in which all relevant information about a scenario, represented by model state $\boldsymbol{u}$, is given. And second, that Pearl gives a unified formal account that can deal both with general and singular causation. Clearly, this feature of a uniform account for both singular and general causation forms a huge advantage of Pearl's account over Hitchcock's.

Let me conclude this introductory section of this chapter with a last brief comparison of Pearl's causal model as defined by (P CM) and Hitchcock's causal model given by (H CM). The immediate observation to be made is that, apart from terminological differences and marginal cases, the formal definition of a causal model is the same for both Hitchcock and Pearl. Hitchcock's notion of a structural equation for a variable Y is in essence just the one of a function f_i from the set of values of the parents of Y to the set of values of Y. In particular, Hitchcock's requirement of asymmetry, demanding that ordinary mathematical operations, like inverting, are not admissible, is met by Pearl's functions f_i.

There is only one marginal difference between Pearl's (P CM) and Hitchcock's (H CM): For Pearl, the system of functions $\boldsymbol{F}$, corresponding to Hitchcock's set of structural equations E, may contain cycles and thus also symmetries, as the marginal case of cycles consisting only of two variables, for certain of its variables. Pearl explicitly discusses a case in which such a symmetry prevails: the equilibrium of supply and demand, in which the set $\boldsymbol{F}$ also has a unique solution, although being cyclic.[91] Hitchcock explicitly also allows cycles, but implicitly rules out

[91] Cf. [jP 00], pp. 215-217.

symmetries.[92] On the other hand, in clause (iii) of his definition (P CM) Pearl explicitly demands that the set of functions F in his causal model has a unique solution, whereas Hitchcock does not mention this property at all.[93]

Also in the same way as for Hitchcock, Pearl's causal models can be graphically represented. This representation follows exactly the same rules, namely drawing an arrow between two variables X and Z, starting at X and pointing to Z, if and only if X is a parent of Z. There, variable X is a parent of variable Z, iff the values of X occur in the argument of f_Z.

Sharing, modulo notation, the same definition of a causal model, the differences between Hitchcock's and Pearl's account lie in the way in which the model is generated and in their definitions of causation, which make use of the model. These were the two problems of causal modeling, as I called them in the introduction of this thesis: the first being the generation of a causal model, and the second being the definition of causation according to a model. I will briefly deal with the first problem in the last section of this chapter, whereas the remainder is devoted to the second problem, the definition of token causation according to a model.

---

[92] Cf. for example [cH 01], p. 281.

[93] On p. 203 of [jP 00], Pearl mentions that if the functions in the set $\boldsymbol{F}$ are determined by recursion, then this ensures that for every realisation $\boldsymbol{u}$ of $\boldsymbol{U}$, the value $\boldsymbol{v}$ of every variable of $\boldsymbol{V}$ is uniquely determined, or in short, that the system of variables has a unique solution. However, uniqueness of solution does not imply recursiveness, as Pearl's discussion of the supply-demand example on p. 280 of [jP 00] shows. For obvious reasons though, the following holds: Recursiveness of $\boldsymbol{F}$ is the property that the system of functions can be ordered in a hierarchy $\boldsymbol{H} = \{f_1, \ldots, f_n\}$, such that for all $k \in \boldsymbol{H}$ the value of function f_k is completely determined by the values of the functions $f_1, \ldots, f_{k-1}$. This property is equivalent to the property that the model does not contain any cycles. Cf. also Pearl's discussion of causal ordering on pp. 227 and 228 of [jP 00].

PEARL'S NOTION OF SUSTENANCE INCORPORATING PRINCIPLES OF PRODUCTION AND DEPENDENCE

In this section, I will discuss the main concept that underlies Pearl's account of token causation, the concept of sustenance. Pearl's basic idea is that for an event *c* to qualify as a cause for another event *e* in a situation *S*, event *c* has to be able to sustain event *e* against certain contingencies. As will turn out in this discussion, these contingencies are structural in nature, which means that they correspond to surgical interventions in a causal model *M* that is employed for formalizing the situation *S* that is to be causally analyzed. Because of this decisive entering of interventions in his conception of sustenance and also because of the counterfactual information encoded in the functions in Pearl's models, his account of token causation can rightfully be classified as both counterfactual and interventionist. In the next sections, I will work out in what way the notion of sustenance enters in Pearl's definitions of actual and contributory causation, his two versions of token causation.

I mentioned in the introduction of this thesis that according to Hall causation has two aspects in our pre-theoretic intuition: production and dependence. I also mentioned that Hall established that the classic counterfactual account of Lewis only captures the dependence aspect and not the production aspect of causation. Pearl in his account of token causation now takes up this view of Hall's and expresses these two aspects in the formal framework of his causal models.

Consider a situation S that contains, among others, events c and e. Suppose that we want to analyze whether c causally produces e and whether e causally depends on c. Let M be a suitable causal model to formalize situation S and suppose in particular that the background variables $\boldsymbol{U}$ of M have the realization $\boldsymbol{u}$. This realization $\boldsymbol{u}$ is called the state of the model M and is intended to formalize the background circumstances that determine the situation S. Furthermore, let X and Y be variables in the causal model M that formalize the possible alterations of the events c and e respectively in situation S. Finally, let $X(\boldsymbol{u}) = x$ and $Y(\boldsymbol{u}) = y$ denote the actual solution for variables X and Y in the model M that results from model state $\boldsymbol{U} = \boldsymbol{u}$, and let these actual values $X(\boldsymbol{u}) = x$ and $Y(\boldsymbol{u}) = y$ formalize the actual events c and e respectively.
If we now for reasons of greater simplicity identify the formal devices in the causal model M with their references in situation S, we yield the following characterization of production and dependence:[94]

Characterization (Production and Dependence, Pearl) **(P PD)**

The *dependence* aspect of causation appeals to the necessity of a cause x in maintaining its effect y in the actual circumstances $\boldsymbol{u}$ in the face of certain contingencies, which otherwise negate y:

Suppose $X(\boldsymbol{u}) = x$ and $Y(\boldsymbol{u}) = y$.

Then y *causally depends* on x if and only if $Y_{x'}(\boldsymbol{u}) = y' \neq y$ for all other values x' of X.

[94] For Pearl's own remarks about production and dependence cf. [jP 00], p. 316. My characterization here deviates from Pearl's original formulation in order to add clarity. Observe in particular that Pearl himself does not even distinguish the formal devices in the model M from their references in the situation S.

The *production* aspect of causation appeals to the capacity of a cause x to bring about the effect y in certain circumstances $\boldsymbol{u}'$, which are counter to fact and in which both x and y were previously be absent:

Suppose $X(\boldsymbol{u}') = x'$ and $Y(\boldsymbol{u}') = y'$
for a certain value $x' \neq x$ of X and certain circumstances $\boldsymbol{u}' \neq \boldsymbol{u}$.
Then x *produces* y if and only if $Y_x(\boldsymbol{u}') = y$.

Clearly, both these aspects of token causation appeal to surgical interventions. In the case of dependence, an intervention on variable X takes place, changing its value from the actual $X(\boldsymbol{u}) = x$ to a value $x' \neq x$ of X that is counter to fact. In the case of production, an intervention then changes the value $x' \neq x$ of X being counter to fact, which was brought about by the different circumstances $\boldsymbol{u}' \neq \boldsymbol{u}$ that were counter to fact, to the actual value x of X.

These interventions are denoted by the expression $Y_x(\boldsymbol{u})$ in the above characterization (P PD). $Y_x(\boldsymbol{u})$ formalizes the potential response of variable Y to the surgical intervention $do(X=x)$ in the model $\boldsymbol{M}$ in state $\boldsymbol{u}$ and is defined via the two following definitions:[95]

Definition (Sub-model and Effect of Action, Pearl) **(P SEA)**

Let $\boldsymbol{M} = \langle \boldsymbol{U}, \boldsymbol{V}, \boldsymbol{F} \rangle$, be a causal model, $\boldsymbol{X}$ a set of variables in $\boldsymbol{V}$, and $\boldsymbol{x}$ a particular realization of $\boldsymbol{X}$. The effect of the surgical intervention $do(\boldsymbol{X} = \boldsymbol{x})$ on the model $\boldsymbol{M}$ is given by the following sub-model $\boldsymbol{M_x}$:

$\boldsymbol{M_x} = \langle \boldsymbol{U}, \boldsymbol{V}, \boldsymbol{F_x} \rangle$, where $\boldsymbol{F_x} = \{f_i : V_i \notin \boldsymbol{X}\}_{i=1,\ldots,n} \cup \{\boldsymbol{X} = \boldsymbol{x}\}$.[96]

---

[95] Cf. [jP 00], pp. 204 and 205.

[96] Pearl does not use the term 'surgical intervention'; instead, he refers to these operations as 'actions'. Both terms can be regarded to be synonymous.

Definition (Potential Response and Counterfactual, Pearl) **(P PRC)**

Let M be a causal model, X and Y two subsets of variables in V. The potential response $Y_x(u)$ of variable Y to action $do(X = x)$ in model M is the solution for Y of the set of equations F_x. $Y_x(u)$ can be interpreted as giving a formal expression to the counterfactual phrase 'the value that Y would have obtained, had X been x'.

In this way, clearly Pearl's $do(X = x)$ operator formalizes the notion of a surgical intervention on a set of variables X. For, observe that in the above sub-model M_x, variables X are lifted from the influence of their old functional mechanisms and placed under the influence of new stipulated quasi-mechanisms that set their values x constant; and obviously, no other mechanisms f_i are perturbed in by this operation. In this way, the values of variables X are changed without creating a logical contradiction, namely by turning to a new model M_x.

If the set of functions F in the causal model M does not contain any cycles, so that no variable Z can be an ancestor and a descendant of variables X at the same time, then also backtracking from this intervention by $do(X = x)$ is excluded, since these value changes cannot propagate to the ancestor variables of X.[97] However, I do not want to go into any details with regard to the problem of backtracking here.

[97] For a more elaborate description of Pearl's formalization of interventions, cf. [jP 00], pp. 70 and 205.

Note in particular that in the above definitions (P SEA) and (P PRC) the mechanisms $\{f_i : V_i \in X\}_{i=1,\ldots,n}$ in the model M are not employed for intervening on the variables in X by $do(X = x)$. Neither the associated functions $\{f_i : V_i \in X\}_{i=1,\ldots,n}$ themselves nor any information that they convey are utilized for the intervention on the set X. On the contrary, the functions $\{f_i : V_i \in X\}_{i=1,\ldots,n}$ are overridden by setting all $V_i \in X$ to constant values, thereby rendering all $V_i \in X$ quasi-exogenous.

A peculiar feature of the above notion of production is that production in a certain model state u applies only to different states u' of the model where both x and y are absent. This raises two problems, as Pearl notes:[98] First, evidence about the actual state u cannot be used in the state u' which is used to determine whether x produces y. And worse, second, the fact that x produces y in state u' being counter to fact cannot explain why the value of Y is y in the actual state u.

For this reason, Pearl defines a new concept, called *sustenance*, which forms the core concept in his account of actual causation. Sustenance captures modified notions of dependence and production, the main difference being in the type of contingencies against which x is expected to protect y: In (P PD), the contingencies were circumstantial, namely evolving from a specific model state u. Sustenance protects y from structural modifications of the model itself, that is interventions, which override the equations for a certain set of endogenous variables W by setting them to constant values w. Pearl's definition is the following:[99]

[98] Cf. [jP 00], pp. 316 and 317.
[99] Cf. [jP 00], pp. 316 and 317.

Definition (Sustenance, Pearl) **(P S)**

Let $W \subseteq V$ be a set of variables with realizations w, w', etc., and let $\boldsymbol{u}$ be an arbitrary realization of the set of background variables $\boldsymbol{U}$. We say that x *causally sustains* y in $\boldsymbol{u}$ relative to contingencies in $\boldsymbol{W}$ if and only if

(i) $X(\boldsymbol{u}) = x$;

(ii) $Y(\boldsymbol{u}) = y$;

(iii) $Y_{xw}(\boldsymbol{u}) = y$ for all w; and

(iv) $Y_{x'w'}(\boldsymbol{u}) = y' \neq y$ for some $x' \neq x$ and some w'.

In this formulation (P S), the sustenance feature is mainly expressed by the last two clauses (iii) and (iv). Clause (iii) requires that x alone is sufficient for maintaining y. In other words, if variable X is set to its actual value x in model state $\boldsymbol{u}$ then, even if W is set to any value w that is different from the actual, variable Y still retains its actual value y. In this way, clause (iii) expresses the remainder of the notion of production, for x produces y in all different value distributions of the model, corresponding to the structural modifications by interventions $do(W = w)$. Clause (iv) rules that x is responsible for sustaining y under these conditions. For, there is at least one setting $W = w'$, in which Y takes a value $y' \neq y$, if X is set to $x' \neq x$. Thus, clause (iv) expresses the remainder of the notion of dependence. Taken together, clauses (iii) and (iv) imply that there is one setting $W = w'$ in which x is both necessary and sufficient for y.[100]

[100] For Pearl's own remarks about the concept of sustenance, cf. [jP 00], p. 317.

Let me end this section with a last comment about interventions. Definitions (P SEA) and (P PRC) illustrate that a causal model M in a certain way generates a whole set of models that consists of the respective models that correspond to the possible settings of the *do* operator. In this way, Pearl's models defined by (P CM) already anticipate all admissible surgical interventions on variables. And although I might emphasize a platitude here, let me stress that the functions f_i (pa_i, u_i) in a causal model M defined by Pearl's (P CM) not only formalize independent mechanisms but also have a counterfactual interpretation because they are defined for all admissible values of parent variables. These two facts, the counterfactual meaning of the functions in a causal model given by (P CM) and the entering of surgical interventions defined by (P SEA) and (P PRC) in Pearl's notion of sustenance, in particular clearly show that Pearl's account of token causation can be classified as being counterfactual as well as interventionist.

In the next section of this chapter I will motivate how this notion of sustenance enters in Pearl's definitions of a causal and a natural beam, the basic building blocks of his definitions of actual and contributory causation, his two versions of token causation. A remark that I can already make at this point is that in the above definitions of sustenance (P S) only endogenous variables $W \subseteq V$ formalize structural contingencies. This distinction between exogenous and endogenous variables especially with regard to he fact that Pearl does not admit interventions on exogenous variables will occupy us for the rest of this thesis.

PEARL'S NOTIONS OF A CAUSAL AND A NATURAL BEAM

After having discussed Pearl's notion of sustenance that incorporates principles of production and dependence, it is time to clarify in precisely what form a cause c has to sustain its effect e in a given situation S with a certain constellation of background conditions. I maintain that this can be most easily understood with the aid of the leading idea that c sustains its effect e via a causal process that links cause c and effect e. In the following, I motivate Pearl's definitions of a causal and a natural beam as the decisive step in reaching a formalization of this notion of a sustaining causal process.

Causal or natural beams $\boldsymbol{M_u}$ are projections of a causal model $\boldsymbol{M}$ on its actual state $\boldsymbol{U = u}$. This state formalizes the constellation of background conditions in the given situation S that is to be analysed for causal dependencies. I will show in this section that a causal or a natural beam $\boldsymbol{M_u}$ describe under which conditions a set of parent variables S_i locally sustains the actual value v_i of their child variable V_i, and this for every endogenous variable V_i in the causal model $\boldsymbol{M}$ in the state $\boldsymbol{u}$.

In the next section I will then clarify what it means for the value x of an arbitrary variable X to sustain a value y of another arbitrary variable Y in a causal model $\boldsymbol{M}$ in the state $\boldsymbol{u}$. These conditions are given by Pearl's definitions of actual and contributory causation. With this, it will then be possible to determine precisely what a sustaining causal process is for Pearl.

I remarked in the last section that Pearl's definition of sustenance (P S) regulates against which kind of contingencies a cause, formally described by $X = x$, must sustain its effect, modeled by $Y = y$. These contingencies are structural in nature, meaning that they are brought about by interventions on a set of variables $W \subseteq V$. What definition (P S) is missing is a characterization of what choices are admissible for these contingency sets W.

Clearly the choice of these sets is crucial. For example, an unrestricted set W could contain a variable Z, that is a descendant of X and ancestor of Y at the same time. Obviously, under these circumstances we cannot intervene on X and Z and set them to values $X = x'$ and $Z = z'$ that are incompatible. Otherwise we would disrupt the functions in the causal model M, thereby distorting the formal representation of the mechanisms at work in the situation to be modeled.[101]

For this reason, namely to specify which choices of contingency sets to be intervened on are admissible, Pearl defines a causal beam. Causal beams are projections of a causal model M to its actual state u, devised to make the choice of contingency sets W minimally disruptive. Pearl's definition is the following:[102]

Definition (Causal Beam, Pearl) **(P CB)**

For causal model $M = \langle U, V, \{f_i\}_{i=1,\ldots,n}\rangle$ and state $U = u$, a *causal beam* is a new model $M_u = \langle u, V, \{f_i^u\}_{i=1,\ldots,n}\rangle$, in which the set of functions f_i^u is constructed from $\{f_i\}_{i=1,\ldots,n}$ as follows:

[101] In particular, we would risk that clause (iv) in the definition of sustenance (P S) could be wrongly fulfilled by this setting, so that we would wrongly analyze $X = x$ as being a cause of $Y = y$.

[102] Cf. [jP 00], pp. 318 and 319.

1. For each variable $V_i \in \boldsymbol{V}$, partition $\boldsymbol{PA}_i$ into two subsets $\boldsymbol{PA}_i = S_i \cup \underline{S}_i$, where S_i is any subset of $\boldsymbol{PA}_i$ satisfying

   $$f_i(S_i(u), \underline{s}_i, u) = f_i(S_i(u), \underline{s}_i', u)$$

   for all realizations $\underline{s}_i$ and $\underline{s}_i'$ of $\underline{S}_i$.

   In words, S_i is any subset of $\boldsymbol{PA}_i$ sufficient to entail the actual value of $V_i(u)$, regardless of how we set the other members of $\boldsymbol{PA}_i$. [103]

2. For each variable $V_i \in \boldsymbol{V}$, find a subset W_i of $\underline{S}_i$, for which there exists some realization $W_i = w$ that renders the function $f_i(s_i, \underline{S}_{iw}(u), u)$ nontrivial in s_i; that is

   $$f_i(s_i', \underline{S}_{iw}(u), u) \neq f_i(s_i, \underline{S}_{iw}(u), u) = V_i(u)$$

   for some realizations s_i' of S_i.

   $\underline{S}_i$ should be chosen not to intersect the sustaining set S_j of any other variable V_j, for $j \neq i$. Likewise, setting $W_i = w$ should not contradict any other setting W_j for $j \neq i$. [104]

3. Replace $f_i(s_i, \underline{s}_i, u)$ by its projection $f_i^u(s_i)$, given by

   $$f_i^u(s_i) = f_i(s_i, \underline{S}_{iw}(u), u).$$

---

[103] Note that such a sustaining set S_i always exists, since it can always be chosen to exhaust the whole endogenous parent set $\boldsymbol{PA}_i$.

[104] In case the sustaining set S_i exhausts the set $\boldsymbol{PA}_i$, its complement, the contingency set $\underline{S}_i$, is obviously empty, and with this also the non-actual contingency subset W_i. In this case, the above clause 2. is trivially fulfilled, since the projection function $f_i^u(s_i)$ is then identical to the original function $f_i(\boldsymbol{pa}_i, u_i)$.
For the same reason that S_i can always be chosen to equal $\boldsymbol{PA}_i$, the requirement of $S_i \cap \underline{S}_j = \varnothing$ for all $j \neq i$ can always be achieved. Since W_i is a subset of $\underline{S}_i$, it can likewise not interfere with any other sustaining set S_j for $j \neq i$. In the same way, interference of W_i with another W_j for $j \neq i$ can also be avoided by again setting $W_i = \varnothing$.
Note finally that in general the existence of a set $W_i \subseteq \underline{S}_i$ rendering the function $f_i(s_i, \underline{S}_{iw}(u), u)$ nontrivial in s_i is not guaranteed. Even if set W_i exhausts set $\underline{S}_i$, the function $f_i(s_i, \underline{S}_i(u), u)$ can still be constant, despite being dependent on s_i. In this case, S_i is not an admissible choice as a sustaining set.
A discussion of an example, in which a resulting constant projection function rules out the choice of a certain sustaining set, can be found in the next section where I elaborately illustrate the construction of a causal beam.

In the above definition (P CB), the information that the actual state of the causal model M is u is used to simplify the original functions f_i in the model M. This simplification is achieved by determining for each endogenous variable V_i which of its endogenous parent variables in PA_i are responsible for sustaining its actual value v_i in this solution brought about by model state $U = u$.

Clause 1. describes the determination of a sustaining set S_i that is a subset of the endogenous parent set PA_i for an arbitrary endogenous variable V_i with its actual value V_i. The set S_i consists of any endogenous variable v_i that sustains the actual value v_i of V_i against certain structural contingencies in the causal model. These structural contingencies are formalized by possible interventions on the variables in the contingency set $\underline{S}_i := PA_i \backslash S_i$, which is the complement of S_i, in the set of endogenous parents of V_i. In this way, clause (iii) in Pearl's definition of sustenance (P S) is put into concrete terms in this definition of a causal beam.

Clause 2. of definition (P CB) requires that the set S_i is responsible for sustaining the actual value v_i of V_i under these conditions. This requirement is implemented by demanding that there is at least one realization $W_i = w$, in which $f_i(s_i', \underline{S}_{iw}(u), u) \neq v_i$ for a certain realization $s_i' \neq s_i$ of S_i. By this, also clause (iv) of Pearl's definition of sustenance (P S) is put into concrete terms in the definition of a causal beam.

Taken together, clauses 1. and 2. of (P CB) imply that there is at least one setting $W_i = w$ of the parent variables of V_i that are contained in the contingency set $\underline{S}_i$ such that the following holds: The actual realization s_i of the set of parent variables of V_i contained in S_i is both necessary and sufficient for bringing about the actual value v_i of V_i.

Clause 3. finally just rules that the original function $f_i(pa_i, u)$ for variable V_i is to be replaced by its projection $f_i^u(s_i)$ on the sustaining set. By this, the exogenous variables U and the variables in the contingency set $\underline{S}_i$ are

excluded from the argument set of the projection $f_i^{u}(s_i)$ and thereby frozen at model state $\boldsymbol{u}$ and the possibly non-actual realization $\underline{S}_{iw}(\boldsymbol{u})$.

Let us examine closer in what way Pearl's definition of a causal beam (P CB) puts his notion of sustenance expressed in (P S) into concrete terms. The definition of sustenance (P S) declared under which conditions a value x of an arbitrary variable X sustains a value y of another arbitrary variable Y against certain structural contingencies. Formally, these conditions were expressed with the aid of interventions on X and on a certain set of endogenous variables $\boldsymbol{W}$. These interventions froze X and $\boldsymbol{W}$ at various realizations x, x' and $\boldsymbol{w}$, $\boldsymbol{w}'$, so that then the potential response of Y to these interventions $Y_{x\boldsymbol{w}}(\boldsymbol{u}) = y$, and $Y_{x'\boldsymbol{w}'}(\boldsymbol{u}) = y' \neq y$ respectively, could be computed. Remarkably, the background variables $\boldsymbol{U}$ were not frozen at their actual values $\boldsymbol{u}$ by an intervention in this procedure.

The definition of a causal beam (P CB) differs in slight respects from this procedure. First of all, the background variables $\boldsymbol{U}$ are frozen at their actual values $\boldsymbol{u}$ by an intervention. Second, the definition of a causal beam distinguishes between two kinds of contingencies: The variables in the non-actual contingency subset $\boldsymbol{W}_i \subseteq \underline{S}_i$ may be frozen by interventions at non-actual values, whereas the variables in $\underline{S}_i \setminus \boldsymbol{W}_i$ may only be frozen at their actual values by intervening on them.

Finally and most important, clauses 1. – 3. of definition (P CB) only regulate under which conditions a certain set of parent variables S_i locally sustains the actual value v_i of their child variable V_i, and this for every endogenous variable V_i. The question what it means for the value x of an arbitrary variable X to sustain a value y of another arbitrary variable Y, is not answered by this yet. It will be answered though by Pearl's definitions of actual and contributory causation, with which I will deal shortly.

Let me discuss the second difference noted above a bit more elaborately. By allowing non-actual realizations $\boldsymbol{w}$ for the variables in $\boldsymbol{W}_i$ in clause 2. of definition (P CB), one obviously allows a much wider range of

contingencies than if one would require that all variables in the contingency set $\underline{S}_i$ be frozen at their actual values $\underline{S}_i(u)$ brought about by model state $U = u$.

Trivially, clauses 1. – 3. of (P CB) are much easier to fulfill with such a non-actual contingency subset $W_i \subseteq \underline{S}_i$ set to a non-actual realization w than if one would require that $W_i = \varnothing$, so that all variables in the contingency set $\underline{S}_i$ were frozen at their actual values $\underline{S}_i(u)$. For, clearly the requirement that the projection function $f_i^u(s_i)$ is non-trivial in its argument s_i is much stronger if it applies to the actual realization in the model $\underline{S}_i(u)$, so that $f_i^u(s_i) = f_i(s_i, \underline{S}_i(u), u)$, than if it could be met by a non-actual realization $\underline{S}_{iw}(u)$, so that $f_i^u(s_i) = f_i(s_i, \underline{S}_{iw}(u), u)$.

Pearl distinguishes the causal beams M_u in which for all endogenous variables V_i the non-actual contingency subsets W_i are empty as being paradigmatic, since they still describe the actual state of affairs in the situation for which causal model M was constructed. He calls these beams natural and gives them the following definition:[105]

Definition (Natural Beam, Pearl) **(P NB)**

A causal beam M_u is said to be *natural*, if condition 2. in the above definition (P CB) is satisfied with $W_i = \varnothing$ for all $V_i \in V$.

It is constructive to spell this definition out in detail, because this condition of empty non-actual contingency subsets W_i simplifies clauses 2. and 3. in the definition of a causal beam considerably. For this reason, in the following I give my own spelled out version of a definition of a natural beam, which is entirely equivalent to Pearl's original definition.

[105] Cf. [jP 00], p. 319.

Definition (Natural Beam, Spelled-Out Pearl) **(SP NB)**

For causal model $M = \langle \boldsymbol{U}, \boldsymbol{V}, \{f_i\}_{i=1,\ldots,n} \rangle$ and state $\boldsymbol{U} = \boldsymbol{u}$, a *natural beam* is a new model $M_u = \langle \boldsymbol{u}, \boldsymbol{V}, \{f_i^u\}_{i=1,\ldots,n} \rangle$, in which the set of functions f_i^u is constructed from $\{f_i\}_{i=1,\ldots,n}$ as follows:

1. For each variable $V_i \in \boldsymbol{V}$, partition $\boldsymbol{PA}_i$ into two subsets $\boldsymbol{PA}_i = S_i \cup \underline{S}_i$, where S_i is any subset of $\boldsymbol{PA}_i$ satisfying

   $$f_i(S_i(\boldsymbol{u}), \underline{s}_i, \boldsymbol{u}) = f_i(S_i(\boldsymbol{u}), \underline{s}_i', \boldsymbol{u})$$

   for all realizations $\underline{s}_i$ and $\underline{s}_i'$ of $\underline{S}_i$.

   In words, S_i is any subset of $\boldsymbol{PA}_i$ sufficient to entail the actual value of $V_i(\boldsymbol{u})$, regardless of how we set the other members of $\boldsymbol{PA}_i$.

2. For each variable $V_i \in \boldsymbol{V}$, $\underline{S}_i$ renders the function $f_i(s_i, \underline{S}_i(\boldsymbol{u}), \boldsymbol{u})$ nontrivial in s_i; that is

   $$f_i(s_i', \underline{S}_i(\boldsymbol{u}), \boldsymbol{u}) \neq f_i(s_i, \underline{S}_i(\boldsymbol{u}), \boldsymbol{u}) = V_i(\boldsymbol{u})$$

   for some realizations s_i' of S_i.

   Here, $\underline{S}_i$ should be chosen not to intersect the sustaining set S_j of any other variable V_j, for $j \neq i$.

3. Replace $f_i(s_i, \underline{s}_i, \boldsymbol{u})$ by its projection $f_i^u(s_i)$, given by

   $$f_i^u(s_i) = f_i(s_i, \underline{S}_i(\boldsymbol{u}), \boldsymbol{u}).$$

In this formulation, it becomes more obvious that a natural beam is formed by freezing all parent variables in $\underline{S}_i = \boldsymbol{PA}_i \backslash S_i$ that are outside the sustaining set S_i at their actual values $\underline{S}_i(\boldsymbol{u})$, thus yielding the projection $f_i^u(s_i) = f_i(s_i, \underline{S}_i(\boldsymbol{u}), \boldsymbol{u})$.

As a final observation that applies both to natural and non-natural causal beams, observe that for different realizations $\boldsymbol{u}$ of the exogenous variables $\boldsymbol{U}$, the form of a causal beam, and also a natural beam, $\boldsymbol{M_u}$ can vary. This is brought about by clause 1. of Pearl's definitions (P CB) and (P NB) respectively. For different values $V_i(\boldsymbol{u})$ the set of endogenous parent variables S_i sufficient for entailing this value of $V_i(\boldsymbol{u})$ may be different, so that the corresponding set of parent variables $\underline{S}_i$ that are frozen are also different. Furthermore, note that also for one and the same realization $\boldsymbol{u}$ there can be more than one causal beam with different sets S_i reaching from minimally sufficient sets for entailing the actual value of $V_i(\boldsymbol{u})$ to the full endogenous parent set $\boldsymbol{PA}_i$ and different subsets W_i of $\underline{S}_i$.

As I already mentioned in the introduction of this section, the definitions of a causal and a natural beam, (P CB) and (P NB) respectively, are the decisive step in reaching a formalization of the notion of a causal process by which a cause c sustains its effect e. And as we observed in the discussion in this section, (P CB) and (P NB) accomplish this by incorporating Pearl's notion of sustenance. More concretely, we saw that a causal or a natural beam $\boldsymbol{M_u}$ describe under which conditions a set of parents S_i locally sustains the actual value v_i of their child variable V_i, and this for every endogenous variable V_i in the causal model $\boldsymbol{M}$ in the state $\boldsymbol{u}$. Pearl's notion of sustenance expressed in (P S) though also gave guidelines on what it means for the value x of an arbitrary variable X to sustain a value y of another arbitrary variable Y. These guidelines are put into concrete terms by Pearl's definitions of actual and contributory causation, with which I will deal in the next section.

Admittedly, Pearl himself does not talk about causal processes elaborately and only marginally mentions that the formalization of such a process is the guiding idea of his whole account of token causation.[106] So, my reconstruction of his account here is slightly biased. I opted for this motivation of Pearl's account, since his definitions employed are rather technical, and it is easy to overlook the guiding ideas behind them. I am convinced that emphasizing that a formalization of a sustaining causal process linking cause and effect is the leading idea behind his account will aid in developing an intuitive understanding of Pearl's rationale.

In the next section, I will discuss in detail what in my view is the precise formal rendering of a causal process by which a cause *c* sustains its effect *e*. In the section after that, I will with the aid of an example illustrate how exactly an application of Pearl's definition of a natural beam functions on the formal level.

[106] For one of Pearl's short remarks about causal processes, cf. for example [jP 00], p. 313.

PEARL'S NOTIONS OF ACTUAL AND CONTRIBUTORY CAUSATION AND THE FORMALIZATION OF A CAUSAL PROCESS

After having discussed Pearl's definitions of a causal and a natural beam, I now turn to Pearl's definitions of actual and of contributory causation, which finally clarify what it means for the value x of an arbitrary variable X to sustain a value y of another arbitrary variable Y in a causal model M in the state $\boldsymbol{u}$. Pearl distinguishes between two kinds of token causation: actual causation and contributory causation.

His definition of actual causation decisively employs the notion of a natural beam, whereas his definition of contributory causation does the same with the notion of a non-natural causal beam. In essence, x is called an actual cause of y in a causal model M in the state $\boldsymbol{u}$, if and only if x sustains y against the following structural contingencies: for all endogenous variables V_i in M all variables in the local contingency sets $\underline{S}_i$ are frozen at their actual values. Analogously, x is called a contributory cause of y in a causal model M in the state $\boldsymbol{u}$, if and only if x sustains y against the following structural contingencies: for at least one endogenous variable V_i in M there is a non-empty local non-actual contingency sub-set $W_i \subseteq \underline{S}_i$, whose variables are frozen at certain non-actual values.

Pearl intends his definition of contributory causation to formalize a weakened notion of actual causation that exclusively applies to cases of symmetric over-determination. However, I argue that such a distinction between two mutually disjoint kinds of token causation is in no way necessary. For this reason, I offer a trivial unified definition of token causation that contains actual causation as a special case.

After this discussion of Pearl's definitions of actual and contributory causation, I will draw two main conclusions in this section:

My first conclusion concerns my distinction, which I made in the introduction of this thesis, namely between metaphysical accounts of causation as such and pragmatic modeling accounts of causation according to a model. I maintain that Pearl's definitions of actual and contributory causation clearly only give a pragmatic account of what token causation according to a given model amounts to. Although Pearl himself does not explicitly point this out, the concept of a causal model decisively and irreducibly enters in these definitions.

My second conclusion concerns the way in which both Pearl's definitions of actual and contributory causation, employing the concepts of a natural and a causal beam, respectively, formalize the notion of a causal process. I argue that a causal process can be understood to be the concatenation of the projection functions f_i^u in a causal or natural beam M_u of exactly the variables V_i that are intermediate between the cause variable X and the effect variable Y.

Let me start my discussion with Pearl's definition of actual causation. This is the following:[107]

Definition (Actual Cause, Pearl) **(P AC)**

An event $X = x$ is said to be an *actual cause* of $Y = y$ in a state $\boldsymbol{u}$ if and only if there exists a natural beam M_u such that:

$$Y_x = y \text{ in } M_u, \qquad \text{(C1)}$$

and

$$Y_{x'} \neq y \text{ in } M_u \text{ for some } x' \neq x. \qquad \text{(C2)}$$

[107] Cf. [jP 00], p. 319.

This definition of an actual cause now regulates under precisely which conditions the value x of an arbitrary variable X sustains the value y of another arbitrary variable Y in the causal model $\boldsymbol{M}$ in the state $\boldsymbol{u}$ against structural contingencies of a special type: These contingencies are given by freezing all variables in the local contingency sets $\underline{S}_i$ at their actual values for all endogenous variables V_i in $\boldsymbol{M}$.[108]

An actual cause x according to this definition (P AC) fulfils all clauses (i) - (iv) in Pearl's definition of sustenance (P S): The first three clauses (i) - (iii) of (P S) are fulfilled by clause (C1) of (P AC), the last clause (iv) of (P S) by clause (C 2) of (P AC). For, the natural beam $\boldsymbol{M}_{\boldsymbol{u}}$ identifies a special contingency setting $\boldsymbol{W} = \boldsymbol{w}'$ in the terminology of (P S), in which x is both necessary and sufficient for y, and in which no endogenous variable in the model $\boldsymbol{M}$ is frozen at a non-actual value. More concretely, this contingency set $\boldsymbol{W}$ of definition (P S) is constructed by iterating the operation of identifying local sustaining sets S_i and their complements $\underline{S}_i$, the local contingency sets, for all endogenous variables V_i in the causal model $\boldsymbol{M}$. Since $\boldsymbol{M}_{\boldsymbol{u}}$ is a natural beam, all local non-actual contingency sub-sets $W_i \subseteq \underline{S}_i$ are empty, so that all variables in the local contingency sets $\underline{S}_i$ are frozen at their actual values. The total contingency set $\boldsymbol{W}$ in definition (P S) finally is the union of all these local contingency sets $\underline{S}_i$, i.e. it is $\boldsymbol{W} = \bigcup_{i \in \{1,\ldots,n\}} \underline{S}_i$.

In other words, the counterfactual test in (C2) in definition (P AC) ensures that the value y of variable Y would not be sustained by some value x' of X, if for all variables V_i in the causal model $\boldsymbol{M}$ in state $\boldsymbol{u}$ all variables outside the sustaining sets S_i are frozen at their actual values $\underline{S}_i(\boldsymbol{u})$. Condition (C1) would in itself not be enough for making x necessary and sufficient for y, for there could be another value x'' of X bringing about $Y_{x''} = y$.[109]

---

[108] For Pearl's own discussion and further remarks about his definition of actual causation, cf. [jP 00], pp. 319 and 320.

[109] The existence of such a value x'' of X would be admissible by clause 2. in the definition of a natural beam (SP NB).

How can we now graphically understand how according to the definitions of actual causation (P AC) and of a natural beam (P NB) an actual cause c sustains its effect e in a given situation S with a certain constellation of background conditions? I mentioned at the beginning of this chapter that the functions f_i in a causal model M formalize the mechanisms that are at work in the given situation S that is to be analyzed for causal dependencies. In a natural beam M_u now, the projection functions f_i^u formalize special reduced mechanisms for this situation S, if this situation is in a certain fixed state, represented by $\boldsymbol{U} = \boldsymbol{u}$.

These mechanisms are reduced in such a way that all influences that are not immediately involved in sustaining the actual constellation in the situation are screened off. Formally, this screening off of influences from mechanisms is achieved by freezing all variables in the local contingency sets $\underline{S}_i$ in the construction of the projection functions f_i^u. The actual constellation in the situation is formalized by the actual value distribution in the model M. However, the variables in the local contingency sets $\underline{S}_i$ are not only arbitrarily frozen, they are frozen at their actual values. And this amounts not only to screening off the influences that are not immediately involved in sustaining the actual constellation in the situation, but also to fixing these influences at their actual condition.

An actual cause c in this way sustains its effect e against the structural contingencies that correspond to a canonical reduction of the mechanisms at work in situation S to the ones that are able to sustain the actual constellation in the situation, while all not immediately relevant influences are screened off and fixed at their actual condition.

More comments about the graphical understanding of Pearl's definitions of token causation can be found at the end of this section, where I describe what the formal rendering of a causal process is in my view. A first

example illustrating this definition of actual causation (P AC) can be found in the next section of this chapter.[110]

One concluding remark about the role of interventions in Pearl's account of actual causation: Definitions (SP NB) and (P AC) involve surgical interventions in three ways in defining actual causation: First, in the construction of the various natural beams $\boldsymbol{M_u}$, interventions are used for freezing the causal model $\boldsymbol{M}$ at its actual state $\boldsymbol{U} = \boldsymbol{u}$. Second, the variables in the local contingency sets $\underline{S}_i$, are frozen by interventions at their actual values for defining the projection functions $f_i^{\boldsymbol{u}}(s_i) = f_i(s_i, \underline{S}_i(\boldsymbol{u}), \boldsymbol{u})$ for the endogenous variables V_i in model $\boldsymbol{M}$. Third, interventions on variable X are used in order to test whether the actual value x of X is necessary and sufficient for the actual value y of Y in the beam $\boldsymbol{M_u}$.

I mentioned in the last section of this chapter that natural beams are special or paradigmatic causal beams, namely causal beams in which for all variables V_i in the causal model $\boldsymbol{M}$ in state $\boldsymbol{u}$ all variables outside the local sustaining sets S_i are frozen at their actual values $\underline{S}_i(\boldsymbol{u})$. Non-natural causal beams contain non-empty local non-actual contingency sub-sets $W_i \subseteq \underline{S}_i$, whose variables are frozen at non-actual values.

What does admitting a freezing of variables in W_i at non-actual values for the construction of projection functions $f_i^{\boldsymbol{u}}(s_i) = f_i(s_i, \underline{S}_{iw}(\boldsymbol{u}), \boldsymbol{u})$ in a non-natural causal beam $\boldsymbol{M_u}$ amount to graphically? Again, the projection functions $f_i^{\boldsymbol{u}}(s_i) = f_i(s_i, \underline{S}_{iw}(\boldsymbol{u}), \boldsymbol{u})$ formalize reduced mechanisms in the situation S to be analyzed for causal dependencies. And again, these mechanisms are reduced in such a way that all influences that are not immediately involved in sustaining the actual constellation in the situation are screened off. However, this time the variables in the local contingency sets $\underline{S}_i$ are not all frozen at their actual values, some of them are frozen at non-actual values, under the proviso that this is compatible with the actual

[110] Further and more interesting examples can be found in the following chapter.

value distribution in the model. And this amounts to screening off the influences that are not immediately involved in sustaining the actual constellation in the situation, while fixing some these influences at conditions that are counter to fact.

If now an event c sustains another event e against the structural contingencies that correspond to a reduction of the mechanisms at work in situation S to the ones that are able to sustain the actual constellation in the situation, while all not immediately relevant influences are screened off with some of them being fixed at conditions that are counter to fact, then Pearl calls this event c a contributory cause of e. His formal definition is the following:[111]

Definition (Contributory Cause, Pearl) **(P CC)**

An event $X = x$ is said to be a *contributory cause* of $Y = y$ in a state $\boldsymbol{u}$ if and only if there exists a causal beam $\boldsymbol{M_u}$, but no natural beam, such that:

$$Y_x = y \text{ in } \boldsymbol{M_u}, \qquad \text{(C1)}$$

and

$$Y_{x'} \neq y \text{ in } \boldsymbol{M_u} \text{ for some } x' \neq x. \qquad \text{(C2)}$$

Pearl defines this notion of contributory causation in order to account for cases of symmetric over-determination.[112] In essence, a contributory cause is a weakened version of an actual cause for these situations, in which an effect e is symmetrically over-determined by at least two other events c. By employing non-natural causal beams for the definition of contributory causation (P CC), Pearl graphically isolates a set of reduced mechanisms with a decisive property: Although these reduced mechanisms bring about the actual constellation in the situation, they violate the original

[111] For Pearl's definition of contributory causation, cf. [jP 00], p. 319.
[112] For Pearl's own remarks about contributory causation, cf. [jP 00], p. 313.

mechanisms, since some influences that are not immediately relevant for bringing about the actual constellation in the situation are fixed at conditions that are counter to fact.

It is exactly this property that allows the screening off of the influence of an event c_1 that in fact, together with another event c_2, symmetrically over-determines an effect e. And this screening off of the influence of c_1 and fixing it at a condition counter to fact in turn allows determining whether event c_2 would sustain effect e in these slightly altered circumstances. If c_2 sustains e in this way, then c_2 is called a contributory cause of e.

In this way, the rationale behind Pearl's definition of actual causation is straightforwardly adapted to cover cases of symmetric over-determination. Technically, contributory causation could be understood to be the wider and more fundamental concept than actual causation. For, the sustenance of an effect e by a cause c is tested for a wider range of circumstances, namely not only for the actual constellation of the situation, but also for constellations that are counter to fact. Nevertheless, I mainly deal with Pearl's definition of actual causation in this chapter, since I regard this to be the simpler and paradigmatic one, and also since it is the more accessible one. Moreover, situations containing a symmetric over-determination of an effect by at least two events seem to be marginal cases of the type of situations in which we are usually interested.

In any case, Pearl does not really distinguish between actual and contributory causation as being two kinds of token causation. His definitions of actual and contributory causation, (P AC) and (P CC) respectively, share the same underlying rationale of an event sustaining another one in certain circumstances. Also on a formal level, definitions (P AC) and (P CC) are completely analogous, the first employing natural beams in exactly the same way as the second employs non-natural causal beams. For this reason, my remarks made above after introducing the notion of actual causation are equally valid for contributory causation.

In the next chapter, I will present two examples in order to illustrate how Pearl's account of contributory causation is practically applied. For the moment, I would like to draw attention to another point, namely that Pearl's way of formulating actual and contributory causation as being mutually exclusive concepts arbitrarily brings about an unnecessary distinction. Simply deleting the phrase 'but no natural beam' from Pearl's definition of contributory causation (P CC) yields an inclusive notion of contributory causation that comprises actual causation as a special case, since as said every natural beam is in particular a causal beam. With this modification, one could rightfully call this an inclusive definition of token causation containing actual causation as a special case:

Definition (Token Cause, Modified Pearl) **(MP TC)**

An event $X = x$ is said to be a token cause of $Y = y$ in a state $\boldsymbol{u}$ if and only if there exists a causal beam $\boldsymbol{M_u}$ such that:

$Y_x = y$ in $\boldsymbol{M_u}$, (C1)

and

$Y_{x'} \neq y$ in $\boldsymbol{M_u}$ for some $x' \neq x$. (C2)

I will return to this topic of a modification of Pearl's account to yield a definition of inclusive token causation in the next chapter, where I propose such a modification for my extension of Pearl's account. In the present context, let me just say that in my view this distinction between actual and contributory causation as being mutually exclusive concepts is a purely terminological issue and not a philosophical problem.

Before I come to my promised conclusion about the way in which Pearl formalizes a causal process, let me make a final comment about Pearl's definitions of (P AC) and (P CC) that might otherwise not be noticed: Both definitions (P AC) and (P CC) make decisive use of the concept of a causal

model. For, literally they regulate under which conditions a value x of a variable X sustains the value y of another variable Y in a given causal model M in state u. In particular, the choice of the causal model M and its state u completely determine the set of admissible causal beams M_u that are tested for conditions (C1) and (C2).

As a trivial, but important, result, definitions (P AC) and (P CC) only define pragmatic notions of what it is for an event c to cause another event e according to a model M that is given beforehand. In the introduction of this thesis I mentioned the great advantage of a pragmatic account of token causation qualified to a model over a metaphysical account of causation as such. In essence, I argued that such a pragmatic account allows splitting up the problem of analyzing causal dependencies into two smaller and more tangible problems: The first is the generation of a suitable causal model for a given situation, and the second is the definition of token causation according to a given model. This thesis is almost exclusively concerned with the second problem.

Pearl himself does not explicitly make this distinction between metaphysical accounts of causation and pragmatic accounts of causal modeling. Moreover, Pearl nowhere mentions that his definitions (P AC) and (P CC) only account for token causation according to a given causal model. His formulation in (P AC) and (P CC) is even misleading, since it uses a mixed terminology. Although (P AC) and (P CC) only talk of values x and y of variables X and Y in a causal model M and sustaining relationships among them, Pearl refers to $X = x$ and $Y = y$ as events, in this way identifying formal devices in a causal model with the very constituents of a situation the former are intended to formalize. I do not know whether Pearl just intended to present his account more suggestively by this rather casual terminology, or whether he really wanted to account for a kind of causation as such, although I strongly assume the former. In any case, I just wanted to emphasize by this

comment that Pearl's account of token causation is a pure account of causal modeling, and I consider this to be a great advantage.

Let me now draw my final conclusion about the way in which both natural and causal beams formalize the notion of a causal process for Pearl. I argue in the following that a causal process can be understood to be the concatenation of the projection functions $f_i^{\boldsymbol{u}}$ in a causal or natural beam $\boldsymbol{M}_{\boldsymbol{u}}$ of exactly the variables V_i that are intermediate between the cause variable X and the effect variable Y.

Let me start with general observations about the way in which Pearl formalizes a causal process. Basically, both a natural and a causal beam $\boldsymbol{M}_{\boldsymbol{u}}$ are a projection of the causal model previously given on its actual state $\boldsymbol{u}$ incorporating the notion of sustenance. In essence, the exogenous variables $\boldsymbol{U}$ are frozen on their actual values $\boldsymbol{u}$, and this information is used to simplify the functions f_i in the original model to yield the projection functions $f_i^{\boldsymbol{u}}$ in the causal or natural beam $\boldsymbol{M}_{\boldsymbol{u}}$.

Here, for each endogenous variable V_i, the projection function $f_i^{\boldsymbol{u}}$ is formed by determining a set of parent variables $S_i \subseteq \boldsymbol{PA}_i$ that can sustain the actual value v_i of variable V_i against certain contingencies. These contingencies have a structural nature, which means that they are brought about by interventions. A contingency set $\underline{S}_i = \boldsymbol{PA}_i \setminus S_i$ that is comprised of the endogenous parents of V_i that are dispensable for sustaining its actual value v_i is formalizing these contingencies. In case of a natural beam defined by (P NB), all variables in $\underline{S}_i$ are frozen at their actual values with the aid of interventions. In case of a causal non-natural beam defined by (P CB), the variables in a subset $W_i \subseteq \underline{S}_i$ can be frozen at arbitrary non-actual values w_i with the aid of interventions. None of these interventions on $W_i \subseteq \underline{S}_i$ or on $\underline{S}_i$ itself affects the value of v_i of V_i, since this is sustained by S_i. This construction of projection functions $f_i^{\boldsymbol{u}}$ that contain only the variables in the sustaining set S_i as arguments is conducted for every

endogenous variable in the previous causal model M to yield the causal or natural beam M_u.

In what way does a causal or natural beam M_u now formalize the notion of a causal process linking cause and effect? This brings us to the definitions of actual and contributory causation, (P AC) and (P CC) respectively. These definitions rule that a value x of variable X is said to be an actual, or a contributory cause respectively, of the value y of variable Y, if and only if there is a natural beam M_u, or a causal beam M_u but no natural beam respectively, such that clauses (C1) and (C2) hold. Clause (C1) now demands that in the actual solution of the beam M_u variable X has the value x and variable Y the value y, and clause (C2) requires that an intervention on variable X setting X to a certain value $x' \neq x$ propagates in the beam M_u in such a way that this changes the value of Y.

Having made these observations, I understand a causal process to be the concatenation of projection functions f_i^u of exactly the variables V_i that are descendants of X and at the same time ancestors of Y in the causal or natural beam M_u. For, first this concatenation of projection functions f_i^u for the intermediate variables V_i is responsible for bringing about the actual value y of Y in the actual solution of M_u that contains the actual value x of X, thereby fulfilling the requirement posed by clause (C1). And second, this concatenation of projections f_i^u for the intermediate variables V_i is also exactly the way in which the intervention on variable X setting $X=x' \neq x$ propagates in the beam M_u so that a change in the value of Y is brought about. In this way, causal or natural beams are the framework in which causal processes operate. More precisely, value x of variable X is an actual, or a contributory cause respectively, of the value y of variable Y, if and only if there is a natural beam M_u, or a causal beam M_u but no natural beam respectively, in which there is a causal process in which x sustains y.

But how can we graphically understand a causal process now? I mentioned already at the beginning of this chapter that the functions f_i in a

causal model M formalize the mechanisms that are at work in the given situation S that is to be analyzed for causal dependencies. In a causal or natural beam M_u now, the projection functions f_i^u formalize the reduced mechanisms that are at work in the given situation S, if this situation is in a certain fixed state. This means the information that the external factors determining the situation are fixed at a certain constellation is formalized by freezing the model state at $\boldsymbol{U} = \boldsymbol{u}$ and letting this information propagate in the model M. This procedure certainly yields exactly one value distribution for the model M, but for every endogenous variable V_i in M there are possibly more than one reduced mechanism f_i^u that can sustain its value v_i in this solution. This explains why there is usually a family of causal or natural beams M_u associated to a model M at a certain state $\boldsymbol{U}$. In this picture, a causal beam can be understood as formalizing a collection of reduced mechanisms at work in the situation S that is able to sustain the actual constellation in S, if the external factors determining S were previously fixed at a certain state.

An event c in this situation S being fixed at a certain state now is an actual or contributory cause of another event e in S fixed at this state, if and only if there is a causal process linking c and e. With this a causal process is finally a concatenation of reduced mechanisms that is able to sustain the actual constellation of events that are intermediate between c and e in S fixed at this state. In particular this means that this concatenation of reduced mechanisms on the one hand brings about event e if event c is given, and on the other hand that this concatenation of reduced mechanisms also brings about an alteration of event e if c is altered in a certain way. In my view, Pearl's formal constructions expressed in this way capture our pre-theoretic intuitions about a causal process linking token cause and token effect in a situation very well.

AN EXEMPLARY CONSTRUCTION OF A NATURAL BEAM

After heaving dealt with Pearl's definitions of actual and contributory causation employing natural and causal beams and having established that beams play the decisive role in formalizing the notion of a causal process, I now illustrate how a natural beam is constructed according to Pearl's definition (P NB). I focus on the discussion of natural beams here, since they are simpler and in a certain way more paradigmatic than non-natural causal beams. Admittedly, natural beams are special causal beams that are developed to analyze actual causation. In this way causal beams describe a wider concept, and certainly my modified definition of inclusive token causation that employs causal beams contains actual causation as a special case. However, Pearl's original definition of contributory causation is not inclusive, and non-natural causal beams are only applied in analyzing cases of symmetric over-determination.

The functioning of Pearl's definition (P AC) of actual causation, and with this also the functioning of his definition of a natural beam (P NB), can be most easily illustrated with the aid of an example, omitting a cover story and only focusing on a causal model *M*.

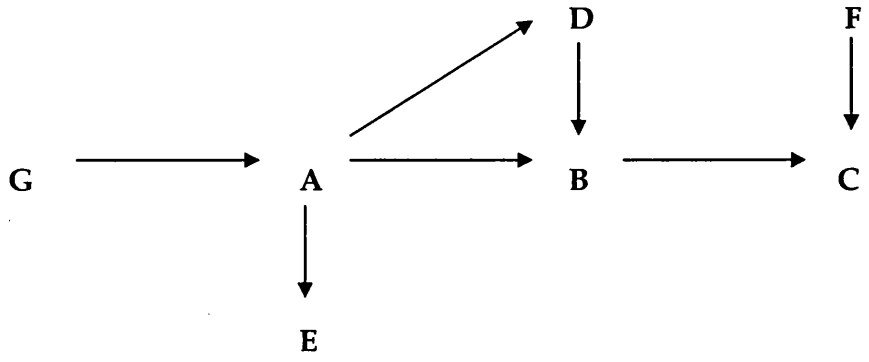
For this purpose I will in the following consider a causal model *M* containing 7 variables, *A*, *B*, *C*, *D*, *E*, *F* and *G*, all of them two-valued, with *F* and *G* exogenous and the following functions for the endogenous variables:

$$f_A(g) = id_G,\ f_D(a) = id_A,\ f_E(a) = id_A,$$

$$f_B(a,d) = \begin{cases} 1, & \text{if } A=1 \text{ and } D=1 \\ 0 & \text{otherwise} \end{cases}, \text{ and } f_C(b,f) = \begin{cases} 1, & \text{if } B=1 \text{ or } F=1 \\ 0 & \text{otherwise} \end{cases}.$$

I further suppose that the actual state of the model is $F = 1$ and $G = 0$, yielding the solution $A = 0$, $E = 0$, $D = 0$, $B = 0$ and $C = 1$.

The associated graph to the model is:



I mainly examine the question whether in this model, M, in the state $F = 1$ and $G = 0$, the actual value $A = 0$ of A is an actual cause of the actual value $C = 1$ of C.

As a first step, I show that for this model M in state $F = 1$ and $G = 0$, there are 3 natural beams, $M_{F=1,G=0}^1$, $M_{F=1,G=0}^2$, and $M_{F=1,G=0}^3$ that accord with definition (SP NB). For, observe that for every model M in state u there are k natural beams M_u^1, ... , M_u^k, with k being the number of possible combinations of sustaining sets for all the variables in the model; the following determination of all sustaining sets for all endogenous variables in M will show that exactly 3 of such combinations exist.

A remarkable observation that I will make in this determination of sustaining sets is that treatment of exogenous variables in constructing sustaining sets is not unequivocally covered by clauses 1. to 3. in Pearl's definition of a natural beam (P NB).

The exogenous variables G and F do not have any sustaining sets at all. Causal beams, and natural beams in particular, are projections of the original model M to its actual state $\boldsymbol{u}$. In this case, the actual state of the model is $F = 1$ and $G = 0$. As a result, F and G are frozen at these actual values.

Variable A has only the exogenous variable G as parent in model M, so that the endogenous parent set $\boldsymbol{PA}_A$ is empty, since according to Pearl's definition (P CM), only endogenous variables can qualify as members of these parent sets. Astonishingly, this case of an endogenous variable that has only exogenous parents is not unequivocally covered by the clauses 1. to 3. in Pearl's definitions of a causal beam (P CB) or a natural beam (P NB), respectively.[113] The reason for my astonishment is that this case of an endogenous variable having only exogenous parents clearly marks the transition from exogenous to endogenous variables in a causal model.
However, I see two straightforward variants for the construction of projection function $f^{G=0,F=1}{}_A$ here: First defining $f^{G=0,F=1}{}_A$ as the constant function $f^{G=0,F=1}{}_A = 0$, embracing the intuition that freezing of the exogenous variables at their actual value should propagate in the beam and hence simplify the projection functions. Second leaving the original function $f_A = \mathrm{id}_G$ in model M unaltered in going over to the beam, so that $f^{G=0,F=1}{}_A = f_A = \mathrm{id}_G$. This second variant is guided by the underlying idea that parent relations between variables should be conserved, so that former endogenous variables are not rendered as being quasi-exogenous, and that projection functions should be non-trivial, in order to allow a propagation of value changes in the beam.

---

[113] I remarked in the last section that natural beams are special causal beams, in which the contingency subsets W are empty. Pearl designed non-natural causal beams to be applied to cases of symmetric over-determination, whereas he took natural beams to apply to all other cases. Since Pearl himself by (P NB) only gave an abbreviated definition of a natural beam, I offered a more elaborate version (SP NB), in which the simplified clauses 1. to 3. with empty contingency subsets W occur.

In my opinion, the second variant is more natural, because it matches Pearl's treatment of endogenous variables that have endogenous parents more harmoniously. For observe that the decisive notion in definitions (P CB) and (P NB) is the notion of sustenance that is expressed by clauses 1. and 2: A projection function should be able to sustain the actual value of an endogenous variable and should moreover be non-trivial. A constant function with no arguments in this way violates both of these conditions. Hence, in our example here, I set the projection function $f^{G=0,F=1}{}_A$ to be identical to the original function $f_A = id_G$ in $\boldsymbol{M}$.

However, observe that one could also make a case for the other variant. One could argue that exogenous and endogenous variables should play a fundamentally different role and that exogenous variables should systematically be excluded from sustaining sets. Pearl definitely has this understanding, at least implicitly. For, in (P CB) and (P NB) he requires the sustaining set S_i for an endogenous variable V_i to be a subset of $\boldsymbol{PA}_i$, the set of endogenous parents of V_i. What Pearl seems to overlook is that this understanding of exogenous variables collides with the endeavor to reach projection functions that sustain the values of variables that were previously endogenous in the original model. In my view, it is the notion of sustenance that fundamentally underlies the construction of a beam, the distinction between exogenous variables and endogenous variables is at best secondary in this respect. In this way, preserving functions that are able to sustain the actual value of endogenous variables should be more important, forcing a strict distinction between exogenous and endogenous variables.

Nevertheless, coming back to our example model here, for Pearl's account of actual causation both variants yield the same result. For, in definition (P AC) Pearl does not allow an intervention on exogenous variables, so that every beam always has the unaltered state $\boldsymbol{U} = \boldsymbol{u}$.

If interventions on exogenous variables were allowed though, there would be a difference between the two variants: For variant one, an intervention on variable G setting $G = 1$ would not propagate through the beam, because $f^{G=0,F=1}{}_A$ is the constant function $f^{G=0,F=1}{}_A = 0$. In variant two, such an intervention would propagate, because $f^{G=0,F=1}{}_A = \mathrm{id}_G$ is clearly dependent of the value of G. Variant one would then appear somehow unnatural and counterintuitive, and it could even be conjectured that the motivation for Pearl not to allow interventions on exogenous variables was mainly to keep this distinction into two variants innocuous.

Variables D and E have variable A as an endogenous parent in the original model M. Hence, clause 1. of definition (P CB) entails that their sustaining sets S_D and S_E equal their parent sets PA_D and PA_E in model M. This renders the projection functions $f^{G=0,F=1}{}_D$ and $f^{G=0,F=1}{}_E$ equal to the original functions $f_D = \mathrm{id}_A$ and $f_E = \mathrm{id}_A$ in M according to clause 2. of definition (P CB). For, observe that in (P CB) the sustaining sets are constructed using the functions f in the original model M.
For variable D, this yields that a choice of $S_D = \varnothing$ as sustaining set and of $\underline{S}_D = \{A\}$ as contingency set is not admissible, since in the original function $f_D = \mathrm{id}_A$, the empty set cannot sustain the value $D = 0$ against the contingency $A = 1$.
Analogously for variable E, setting $S_E = \{A\}$ as sustaining set and of $\underline{S}_E = \varnothing$ as contingency set is the only admissible choice. Obviously, the projection functions $f^{G=0,F=1}{}_D$ and $f^{G=0,F=1}{}_E$ constructed in this way are non-trivial in their arguments according to clause 2. of definition (P CB), because the original functions $f_D = \mathrm{id}_A$ and $f_E = \mathrm{id}_A$ are non-trivial in A.

For variable B, there are three possible choices of sustaining sets, $\{D\}$, $\{A\}$ and $\{A,D\}$, for obviously $A = 0$ as well as $D = 0$ and also the conjunction $A = 0$ & $D = 0$ can sustain the actual value of B, namely $B = 0$.

For variable *C*, we have a similar situation as for variable *A*, because *C* has one exogenous and one endogenous parent. Practically, both the sets {B} and ∅ could sustain the actual value $C = 1$ of *C*, because the exogenous variable *F* is frozen at value $F = 1$. However, in my understanding, the set $S_C = \emptyset$ is not admissible as a sustaining set according to Pearl's definition (P CB), because the resulting projection function would be the constant function $f^{G=0,F=1}{}_C = 1$ and hence be trivial. Apparently, also this case is not unequivocally covered by Pearl's definition of a causal beam (P CB), because this constant function $f^{G=0,F=1}{}_C = 1$ does not have any arguments, and Pearl does not clarify whether such a function has to be understood as a marginal case of a trivial function or not.

However, with the same underlying rationale as in the case of variable *A*, namely preserving parent relationships and not rendering former endogenous variables as quasi-exogenous, I understand the choice of $S_C = \emptyset$ as being unnatural and not homogenously fitting to Pearl's general treatment of endogenous variables. In this way, there is only one choice for the sustaining set of variable *C* left, namely $S_C = \{B\}$.

Note though that again for Pearl's account of actual causation with definition (P AC), both choices of sustaining sets yield the same results. For, a value change of a predecessor variable brought about by an intervention cannot propagate through the beam via variable *C* by either of the resulting projection functions f_C, as the following discussion will show. As an overall observation, we can note that there is an ambiguity in the treatment of exogenous variables in constructing causal beams.

The above discussion shows that there are 3 possible combinations of sustaining sets for the variables in model M with state $F = 1$ and $G = 0$: For, variable B there are three possible choices of sustaining sets, for all other endogenous variables there is only one choice each. It follows that there are 3 natural beams to consider, and one can show that in none of them an intervention on variable A setting $A = 1$ brings about $C = 0$. In accordance with definition (P AC), this means that $A = 0$ is also no actual cause of $C = 1$ in Pearl's account.

As an illustration of how the projection functions in natural beams are computed, construct the natural beam $M_{F=1,G=0^3}$, with the sustaining set $S_B = \{A,D\}$ for variable B.

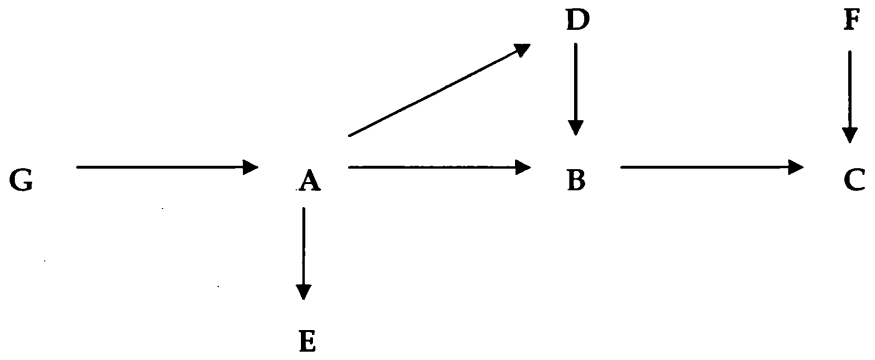
As already mentioned, the exogenous variables F and G are frozen at their actual values, $F = 1$ and $G = 0$, and the projection function $f^{G=0,F=1}{}_A$ for variable A is identical to the original function in M, i.e. $f^{G=0,F=1}{}_A = f_A = \mathrm{id}_G$, yielding the solution $A = 0$.

For variables D and E, the projection functions $f^{G=0,F=1}{}_D = \mathrm{id}_A$ and $f^{G=0,F=1}{}_E = \mathrm{id}_A$ have also been established above, so that the solution for these variables is $D = 0$ and $E = 0$.

Since for variable B, the choice of sustaining set $S_B = \{A,D\}$ and complement $\underline{S}_B = \varnothing$ again exhausts the parent set PA_B of B in M, the projection function $f^{G=0,F=1}{}_B$ is once more identical to the original function f_B in M with $f_B = \begin{cases} 1, & \text{if } A = 1 \text{ and } D = 1 \\ 0 & \text{otherwise} \end{cases}$, resulting in the solution $B = 0$.

The same finally applies to variable C. The sustaining set $S_C = \{B\}$ and contingency set $\underline{S}_C = \varnothing$ result in the projection function $f^{G=0,F=1}{}_C$ being identical to the original function $f_C = \begin{cases} 1, & \text{if } B = 1 \text{ or } F = 1 \\ 0 & \text{otherwise} \end{cases}$.

The graph associated to this natural beam $M_{F=1,G=0^3}$ is hence still the following one:



Returning to our short discussion of the choice of sustaining sets for variable C, note that in all three beams $M_{F=1,G=0^1}$, $M_{F=1,G=0^2}$, $M_{F=1,G=0^3}$ with sustaining set $S_C = \{B\}$ the projection function $f^{G=0,F=1}{}_C$ gives the same value $f^{G=0,F=1}{}_C = 1$, as long as the value of the exogenous variable F is preserved. For trivial reasons the same would hold for the beams $M_{F=1,G=0^4}$, $M_{F=1,G=0^5}$, $M_{F=1,G=0^6}$ with sustaining set $S_C = \varnothing$, if such a choice was admissible, since there we would have the constant function value $f^{G=0,F=1}{}_C = 1$.
As a result, the claim from above is validated, namely that in none of the three beams for model state $G = 0$, $F = 1$ an intervention on variable A setting $A = 1$ brings about a change in the value of variable C, so that $A = 0$ is not an actual cause for $C = 1$ according to Pearl.

Finally note that he number and form of the natural beams to be considered can vary significantly between different value distributions in the causal model M. E.g. for the state $G = 1$, $F = 0$, and the resulting solution $A = 1$, $E = 1$, $D = 1$, $B = 1$ and $C = 1$, we have only one natural beam, since there is only one sustaining set for variable B, namely {A,D}. Observe that this beam $M_{G=1,F=0}$ has different features from the above three beams for state $G = 0$, $F = 1$. Although $M_{G=1,F=0}$ still has the same sustaining set $S^{G=1,F=0}{}_C = \{B\}$ for variable C, this time an intervention on variable A setting $A = 0$ would propagate through the beam, thereby changing the value of C to $C = 0$.

A Comparison of the Formalizations of a Causal Process by Pearl and Hitchcock and Pearl's Problem With Exogenous Variables as Causes

In the last section, I have illustrated how a natural beam according to Pearl's definition (P NB) is constructed for a given causal model. I settled for an illustration of the construction of a natural beam, since I considered natural beams in a certain way to be simpler and more paradigmatic than causal beams. In the section before that, I argued that a causal process for Pearl can be understood to be the concatenation of the projection functions f_i^u in a causal or natural beam M_u of exactly the variables V_i that are intermediate between the cause variable X and the effect variable Y. In this section, I point out striking differences between Pearl's formalization of a causal process with the aid of natural and causal beams on the one hand and Hitchcock's formalization of a causal process by a simple causal route on the other. Finally, I draw attention to the fact that values of exogenous variables can neither be actual nor contributory causes for Pearl, mainly since he excludes exogenous variables from sustaining sets in the construction of both a natural and a causal beam.

After having worked out the motivation behind Pearl's natural and causal beams and the way in which they formalize the idea of a causal process in a previous section, let me now briefly point out some formal differences between a causal process for Pearl and Hitchcock's formalization of a causal process by a route in a model. I will deal with Hitchcock's methodology of formally identifying a causal process with a route in greater detail in the next chapter, where I also claim that this identification

is responsible for a failure of Hitchcock's account with regard to examples of a certain type.

As became obvious, both Pearl and Hitchcock share the conviction that a causal process in a situation S should link cause and effect via the intermediate events, or in the formal setting, that cause variable X and effect variable Y should be linked by variables intermediate between them. Hitchcock and Pearl just differ in the way in which these intermediate variables are selected and in which variables are frozen in this procedure.

Hitchcock formally identifies causal processes with routes r between X and Y in the causal model and freezes intermediate variables between X and Y that do not belong to r at their actual values in order to give an account of actual causation. In order to account for contributory causation, Hitchcock allows a freezing of variables W at arbitrary values independently from whether they are intermediate between X and Y or not. The only conditions are that these variables W do not belong to r and that their new values do not affect the actual value of Y.

Pearl's proceeding differs from this mainly by first freezing the exogenous variables to simplify the given situation and then freezing the variables $\underline{S}_i$ at certain values that are not involved in sustaining the actual values v_i of the endogenous variables V_i in the model. Whether these variables $\underline{S}_i$ are all frozen at their actual values or whether a certain subset $W \subseteq \underline{S}_i$ is frozen at non-actual values depends on whether Pearl wants to account for actual or contributory causation. In any case only the variables V_i matter that are intermediate between X and Y. This difference between Pearl's and Hitchcock's strategies in selecting intermediate variables and freezing others results in a divergence of their formalizations of a causal process in mainly two respects:

First, for Pearl a causal process is dependent on the actual values v_i of the endogenous variables V_i being intermediate between X and Y that are to be sustained. Since a causal model is deterministic, though, these values v_i of the intermediate V_i are determined by the actual state $\boldsymbol{U}$ at which the model is frozen. We saw in the discussion of the formal causal model in the last section that for model state G = 1 and F = 0 there existed only one natural beam $\boldsymbol{M}_{G=1,F=0}$, whereas state $G = 0$ and $F = 1$ had three associated beams $\boldsymbol{M}_{G=0,F=1}$. In contrast to this, Hitchcock's causal routes are independent from the value distribution in a model; the only constraint for the selection of intermediate variables V_i is that they are descendants of X and ancestors of Y.

Second, a process in Pearl's understanding can contain several routes. For, as we saw this, too, in the discussion of the formal causal model in the last section, one sustaining set for variable B contained two variables, namely A and D. In this way, the intermediate variables between A and C that were responsible for sustaining the actual value 1 of C were B and D, and A was a parent variable of both B and D. This means the set of intermediate variables responsible for sustaining C at its value $C = 1$ contained the two routes (A, B, C) and (A, D, B, C). Admittedly, we had no real causal process here linking A and C, since $A = 0$ was not an actual cause for $C = 1$ according to Pearl. However, in the next chapter we will encounter two examples in which we have proper causal processes linking cause and effect and containing two routes.

As a close of this discussion of causal processes and for reasons of instructiveness, I finally apply Hitchcock's account to the causal model given in the last section. Hitchcock despite having a static criterion of identifying causal processes with routes yields the same result as an application of Pearl's account: In our original model $\boldsymbol{M}$ in state $G = 0$, $F = 1$ an intervention on A setting $A = 1$ does not alter the value of C. In this way, $A = 0$ is not an actual cause of $C = 1$ for Hitchcock either.

Hitchcock's definition (H AC) rules that $A = 0$ is a cause of $C = 1$, if and only if one of the two causal routes, (A,B,C) and (A,D,B,C), leading from A to C are active. Verify with the aid of Hitchcock's definition (H AR) that both (A,B,C) and (A,D,B,C) are inactive:

Route (A,B,C) is active iff in the associated dual model ${}_{(A,B,C)}M'$, which is identical to the original model M with the sole exception of the equation for variable D, intervening on A setting $A = 1$ yields $C = 0$. Since an intervention on A only affects descendants of A, and of these only the ones are relevant that are ancestors of C at the same time, all equations apart from the one for variable D remain unaltered in the transition from M to ${}_{(A,B,C)}M'$. Only variable D is intermediate between A and C in this respect, being disjoint from route (A,B,C), so that only its equation is overridden by $D = 0$, thereby freezing D at its actual value. Apparently, then an intervention setting $A = 1$ in the dual model ${}_{(A,B,C)}M'$ still brings about $C = 1$. This shows that route (A,B,C) is not active.

For route (A,D,B,C), the dual model ${}_{(A,D,B,C)}M'$ is identical to the original model M, since there are no intermediate variables between A and C that do not belong to route (A,D,B,C). Also in this model, an intervention setting $A = 1$ yields $C = 1$. This shows that route (A,D,B,C) is not active either. As a result, according to Hitchcock, $A = 0$ is not a cause of $C = 1$. In this way, the analysis of this example by Hitchcock's account of actual causation with definitions (H AR) and (H AC) yields the same verdict as an application of Pearl's account of actual causation with definitions (P NB) and (P AC).

This example conclusively illustrates that whether in the transition in Hitchcock's account from M to ${}_{r}M'$ for a certain route r a certain variable X is frozen at its actual value depends only on whether X is intermediate between the two variables in question. This criterion of being intermediate is static, i.e. independent of the concrete distribution of values of the variables in the model, always the same variables are frozen. More

formally, for Hitchcock, the dual models M' are the same for all possible states of the model, modulo the value at which the intermediate variables are frozen. For example, for state $F = 0$, $G = 1$, model ${}_{(A,D,B,C)}M_{F=0,G=1}'$ is still identical to the original model M, and ${}_{(A,B,C)}M_{F=0,G=1}'$ is identical to the above ${}_{(A,B,C)}M'$, with the only difference being that variable D is frozen at $D = 1$.
As seen above, this rationale of Hitchcock's to identify a causal process with a route in a causal model differs in two important ways from Pearl's methodology to formalize a process by a causal beam. First, the projection functions in a causal beam, and with this the causal beam itself, are dependent on the actual values in the model. And second, projection functions can have more than one argument.

After having made this comparison of Pearl's and Hitchcock's ways to formalize causal processes, let me turn to another topic. We already noticed in the last section of this chapter that clauses 1. to 3. in Pearl's definitions of a causal and a natural beam, (P CB) and (P NB) respectively, contain an ambiguity in the treatment of exogenous variables in forming certain sustaining sets. However, this ambiguity is part of a bigger problem regarding the general role of exogenous variables.
In the last section I argued for including exogenous variables in sustaining sets for endogenous variables if this is the only way to reach a non-trivial projection function that would otherwise have no arguments and be constant. My argument for this was that I regarded the notion of sustenance to be more fundamental than a strict distinction between the roles of exogenous and endogenous variables. However, it became clear that in this way Pearl's rationale of attributing a different role to exogenous variables than to endogenous variables would violated. For, Pearl constructs his causal and natural beams in such a way that in general exogenous variables are to be excluded from sustaining sets and in this way do not enter as arguments of the projection functions. This general

formulation of a causal beam though results in the fact that changes in the values of exogenous variables brought about by interventions, if such interventions were allowed, cannot propagate in the beam. In this way, values of exogenous variables can neither be actual nor contributory causes for Pearl.

However, this exclusion of exogenous variables from sustaining sets in general is only the most striking difference that Pearl makes in treating exogenous and endogenous variables. All in all, he differentiates the role of exogenous variables from the one of endogenous variables in three respects. First, observe that in the definition of a causal model (P CM), the third clause (iii) differentiates between the endogenous parents $\boldsymbol{PA}_i$ and the exogenous parents $\boldsymbol{U}_i$ of an endogenous variable V_i, by representing the value v_i of V_i by the function f_i $(\boldsymbol{pa}_i, \boldsymbol{u}_i)$.

Second and most important, as already mentioned, in the definitions of a causal beam (P CB) and a natural beam (P NB) exogenous variables are excluded from the construction of sustaining sets. Already the clause 1. rules that the sustaining set S_i of the endogenous variable V_i has to be a subset of $\boldsymbol{PA}_i$, already assuming that all exogenous variables $\boldsymbol{U}$, and in this way in particular the exogenous parents $\boldsymbol{U}_i$ of V_i, are frozen at their actual values. Clauses 2. and 3. then ensure a non-triviality of the projection function $f_i^u(s_i)$ constructed for V_i in this way, and replace the original function f_i $(\boldsymbol{pa}_i, \boldsymbol{u}_i)$ with this projection $f_i^u(s_i)$ that has significantly fewer arguments. In this way, even if an intervention on exogenous variables were admissible, it could not propagate through the causal beam, since exogenous variables are no longer arguments of the projection functions $f_i^u(s_i)$.

Third and more a trivial notational issue, the definitions of actual and contributory causation, (P AC) and (P CC) respectively, in their given formulation seem to prohibit an intervention on exogenous variables. For, observe that the counterfactual test by clause (C2) still applies to the causal

beam M_u, in which all exogenous variables $\boldsymbol{U}$ are frozen at their actual values u, so that the variable X intervened on can only be exogenous.

Pearl neither mentions that exogenous variables can neither be actual nor contributory causes, nor does he give a clear rationale for not admitting interventions on exogenous variables. An underlying idea can though easily be conjectured: Pearl understands the values of exogenous variables, or background variables, as he calls them, to be determined by factors outside the model. In contrast to this, the derivation of the values of the endogenous variables is completely known. It is the mechanisms in the example to be modelled reflected by the set of functions F in the model that determine the values of endogenous variables. One could argue that a manipulability of the value of a variable requires knowledge of a mechanism by which this value can be manipulated. In this way, if the laws governing the values of variable are unknown, then these variables are beyond our control, i.e. an intervention on them is not admissible for conceptual reasons.

I argue though such a view is not justified, for the following rather trivial reason: Pearl himself does not employ the mechanisms in the model for intervening on a certain variable X. The associated function f_X for variable X is in no way utilized for the intervention on X, nor is any information that f_X conveys utilized for this purpose. On the contrary, the function f_X is overridden by setting the value of X to a constant value, thereby rendering X quasi-exogenous.

In the next chapter, I will propose an extension of Pearl's account that treats all variables uniformly and allows an intervention on exogenous variables. It will show that it can analyze several examples much more naturally than Pearl's original approach.

PEARL'S PRAGMATIC STRATEGY TO GENERATE BASIC CAUSAL STRUCTURES OUT OF PROBABILITY DISTRIBUTIONS

In this section, I will give a brief synopsis of how Pearl generates his causal models and compare this to the way in which Hitchcock arrives at his causal models for a given example situation. Pearl's approach to generating causal models is mainly geared to analyzing causal relationships on the type level. Pearl's starting point is the definition of conditional probabilistic independence for a set of random variables $T := U \cup V$ with a joint probability distribution P extending over all admissible, in particular over all non-actual combinations of values of the variables in $U \cup V$. Pearl utilizes this probability distribution P and this definition of conditional probabilistic independence in his so-called algorithm of inductive causation to determine an equivalence class $[D]$, a so-called pattern, of basic causal structures D that can all generate this probability distribution P. In this way, the result of Pearl's model generating procedure is not a complete causal model M with a set of deterministic functions F, but instead, the result is the equivalence class $[D]$ of basic causal structures D that are in essence just the set of variables $U \cup V$ together with a relation R describing parenthood between these variables.

Admittedly, my synopsis is in no way complete nor is it self-contained. Instead, I would like to refer the reader to Pearl's own discussion of the model generation procedure, which is far more thorough and comprehensive. I mention Pearl's model generation procedure merely for the following reasons:

First, for drawing the conclusion that in the application of Pearl's definitions of actual or contributory causation, the causal model M out of which the beams M_u are constructed is strongly underdetermined, because the set of functions F determining the values of the variables in V has to be stipulated. Second, to compare Pearl's approach with the one of Hitchcock.

Apart from this, the only critical remark that I will have to make regarding Pearl's account of model generation is that his stipulation of a probability distribution P over a set of random variables $\boldsymbol{U} \cup V$ that formalize possible alterations of *singular* events could be problematic and is definitely in need of motivation.

In comparison with Hitchcock's account of model generation, the following striking difference will become apparent: Pearl has a strictly pragmatic strategy employing probability distributions to extract causal patterns out of them and ending up with a multitude of basic causal structures that can all serve as blueprints for a complete causal model. Hitchcock's strategy is more metaphysical in nature because the complete counterfactual information about a situation is stipulated at the outset of model construction and uniquely and completely determines a causal model on which Hitchcock's definitions of actual or contributory causation are applied.

In this way, the advantage of a causal modeling account over classic metaphysical accounts of causation, namely to divide the analysis of causation into two sub-problems – first the generation of a suitable causal model to formalize a given situation and second the definition of causation qualified to such a model – is lost. As a result of this comparison, I will finally maintain that Hitchcock fails in his attempt to simplify Pearl's account of the first problem of causal modeling, namely the generation of a suitable causal model to formalize a given situation S.

Pearl's account of the generation of a causal model addresses the question, how we can translate direct observations into causal relationships. His main interest is to give a basic rationale for constructing causal models analyzing causation on the type level. For this reason, Pearl is mainly concerned with the problem of deriving causal relationships from statistical data, which give information about the covariation between certain events or types of events. Pearl describes the procedure of generating a causal model as being in essence an adaptation of the procedure of scientific induction to extract causal patterns out of probabilistic relationships.[114]

In the following, I very briefly sketch how in my opinion this procedure of extracting causal patterns out of statistical relationships can most easily be understood when it is applied to analyze causal dependencies on the token level between singular events in a given situation *S*. As already mentioned, my sketch is in no way complete nor is it self-contained. For a more thorough and comprehensive description, I would hence like to refer the reader to Pearl's own discussion of the model generation procedure.[115]

In my understanding, the starting point of Pearl's model generation procedure has to be a set $\boldsymbol{T} := \boldsymbol{U} \cup \boldsymbol{V}$ of random variables formalizing the events and their relevant alterations in situation *S*. The statistical data that is available about this situation is given in the form of a joint probability distribution P over the set $\boldsymbol{U} \cup \boldsymbol{V}$. This distribution P assigns to every possible combination of values $(u_1,\ldots,u_m,v_1,\ldots,v_n)$ for all of the variables $(U_1,\ldots,U_m) \in \boldsymbol{U}$ and $(V_1,\ldots,V_n) \in \boldsymbol{V}$ a probability $p(u_1,\ldots,u_m,v_1,\ldots,v_n) \in [0,1]$.

---

[114] For an extended motivation of Pearl's views of the model generating procedure cf. pp. 41-43 and also pp. 59-64 of [jP 00].

[115] For Pearl's discussion of extracting basic causal structures out of probabilistic information, cf. pp. 43-54 of [jP 00]. For the causal interpretation of these basic causal structures cf. pp. 54-57 of [jP 00].

Note in particular that only one of these combinations $(u_1,\ldots,u_m,v_1,\ldots,v_n)$ of values for the variables in $U\cup V$ gives their actual values $u_1^0,\ldots,u_m^0,v_1^0,\ldots,v_n^0$, all the other combinations $(u_1,\ldots,u_m,v_1,\ldots,v_n)$ are counter to fact.

Pearl's strategy is first to determine the conditional probabilistic independencies in this set $U\cup V$ for this distribution P. He achieves this by the following definition:[116]

Definition (Conditional Probabilistic Independence, Pearl) **(P CPI)**

Let P be a joint probability distribution over a set $T := U\cup V$ of variables, and let X, Y, Z be any three subsets of T.

The set X is said to be *conditionally independent of* Y given Z, written $(X \perp\!\!\!\perp Y \mid Z)_P$, if for all realization s x of X and all realizations y of Y and z of Z with $P(Y=y, Z=z) > 0$ the following holds:

$$P(X=x \mid Y=y, Z=z) = P(X=x \mid Z=z).$$

In words, learning the realization of set Y does not provide additional information about the set X, once we know the realization of set Z. In my understanding, it is these conditional probabilistic independencies in set $U\cup V$ according to probability distribution P out of which Pearl tries to extract causal information. In essence Pearl tries to condense a so-called basic causal structure out of the statistical data given by probability distribution P. Pearl's leading idea for this extraction of causal relationships out of probabilistic data is conveyed by the following definition of inferred causation:[117]

---

[116] Cf. [jP 00], p. 11.

[117] Cf. [jP 00], pp. 45 and 46 for Pearl's original version. My formulation here slightly deviates from Pearl's original, sine Pearl defined inferred causation for so-called latent structures containing latent variables that cannot be observed in the data and have to be stipulated rather than extracted.

Definition (Inferred Causation, Pearl) **(P IC)**

> Given a probability distribution P over a set $V \cup U$ of variables, a variable C has a *causal influence* on variable E if and only if there exists a directed path from C to E in every minimal basic causal structure D that is consistent with P.

In this context, a basic causal structure is defined as follows:[118]

Definition (Basic Causal Structure, Pearl) **(P BCS)**

> A basic causal structure D of a set of variables $T := U \cup V$ is a directed acyclic graph in which each node corresponds to a distinct element of $U \cup V$, and each link represents a direct functional relationship among the corresponding variables.

Pearl does not clarify in which this direct functional relationship consists. However, I maintain that prima facie it can just be understood as expressing a naïve causal relation R of parenthood, or direct causal dependence, between variables. In broad terms, R holds between two variables Z and Y, written $R(Z,Y)$, if and only if Y is directly causally dependent on Z in pre-theoretic understanding. The relation R induces a natural distinction between exogenous and endogenous variables U and V. The only formal requirement that R has to meet is that it orders the variables in $U \cup V$ in a directed acyclic way. Let me briefly record this understanding of mine of a basic causal structure by the following characterization:

---

[118] Cf. [jP 00], p. 44 for Pearl's original definition. Pearl himself does not call his causal structures 'basic'. This is my addition in order to add clarity to the distinction between these structures D and complete causal models M.

Characterization (Basic Causal Structure, Extended Pearl) **(EP BCS)**

A basic causal structure D is a pair $D = \langle T, R \rangle$ of a set of variables $T := U \cup V$ and a dyadic directed acyclic relation R that expresses a naive notion of causal parenthood relative to T.

In order to render his definition of inferred causation (P IC) operational, Pearl has to clarify what it means for a basic causal structure to be minimal. According to Pearl, his notion of minimality is in essence a formalization of Occam's razor, the standard norm of scientific induction. However, I do not intend to argue for Pearl's notion of minimality here. Instead, I would like to refer the reader to Pearl's own argumentation, in order to form an independent opinion for him- or herself.[119]

Pearl implements this definition of inferred causation (P IC) in his algorithm of inferred causation.[120] He develops this algorithm with the target to reconstruct the structure of a directed acyclic graph from the given probability distribution P via queries about the conditional independencies that are embodied in P according to his definition (P CPI). More precisely, this algorithm is designed to give a graphical representation of the equivalence class $[D]$ of some minimal basic causal structure D that can generate P.[121] Pearl calls this representation of $[D]$ a *pattern H* and characterizes it as a partially directed acyclic graph, in which some edges are directed and some are undirected. Again, I do not intend to argue for Pearl's algorithm of inductive causation. Instead, would like to refer the reader again to Pearl's own argumentation, in order to form an independent opinion for him- or herself.[122]

---

[119] Cf. mainly pp. 45-48 of [jP 00] for this.

[120] Cf. pp. 49-51 of [jP 00] for Pearl's algorithm of inductive causation.

[121] Pearl elaborately argues that a minimal basic causal structure that is able to generate a probability distribution P of a certain type is only determined up to independency equivalence. Cf. [jP 00], pp. 48-51 and p. 19 for this argumentation.

[122] This can be found on pp. 49-51. Admittedly, Pearl also discusses a weakened version of his algorithm of inductive causation that can deal with latent variables and produces a

The lesson I would like to draw from this brief exposition of Pearl's way of generating causal models is the following: The result of Pearl's model generating procedure is not a complete causal model M with a set of deterministic functions F. Instead, Pearl's algorithm of inductive causation only produces an equivalence class $[D]$ of basic causal structures D. And such a basic causal structure is in essence just the set of variables $\boldsymbol{U} \cup \boldsymbol{V}$ together with a relation R describing a naïve causal notion of parenthood between these variables.

However, as seen in this chapter, a complete causal model M is required in order to apply Pearl's definitions of actual and contributory causation, (P AC) and (P CC) respectively. It is the set of deterministic functions F that is the decisive component in a causal model for this purpose. For, these functions f_i do not only describe the causal mechanisms that are at work in the situation to be modeled, but they also determine the natural and causal beams M_u that can be constructed out of the original model M. Such a set of functions F has to be stipulated, if we extend a basic causal structure D to a complete causal model. The causal parent relation R contained in D surely delivers constraints on the admissible choice of F, but it still leaves this set highly under-determined.

In short, Pearl's model generating procedure is a sophisticated and highly interesting way to derive causal parent relationships from statistical data. However, this procedure has the severe shortcoming that it does not give us any information, where the deterministic functions f_i come from, without which we cannot analyze token causation according to Pearl.

As a last technical remark before I come to the final comparison between Pearl's strategy to construct causal models and Hitchcock's procedure, let me note the following: Pearl's algorithm of inductive causation only produces an equivalence class $[D]$ of basic causal structures D, and the

---

weaker result than a pattern [D]. For details about this version of the algorithm cf. pp. 51-54 of [jP 00].

latter have the form of a directed acyclic graph. However, according to Pearl's definition (P CM) of a complete causal model the set of deterministic functions F does not have to be acyclic. (P CM) only requires that F has a unique solution. This divergence between a basic causal structure and a complete causal model could also be understood as suggesting that Pearl's model generating procedure is still in need of completion. But be this as it may, the results of this thesis are not affected by whether Pearl might in the end want his causal models to be acyclic or not. All examples that I employ during my argumentation in the next chapter involve sets of functions F that have a unique solution and do not contain any cycles.

Let me now finally compare Pearl's strategy for generating causal models with Hitchcock's procedure. A rather trivial difference between Pearl's and Hitchcock's accounts of model generation is the following: Hitchcock attempted to describe how we arrive from a given description of an arbitrary example situation S at a causal model M for this situation. The essential steps were: a) Establishment of the relevant events and their possible alterations; and establishment of corresponding variables and assignments of their values to event-alterations. b) Determination of a basic set of counterfactuals. c) Formalization of this given counterfactual information in a set of structural equations.
The first of these steps – establishment of the relevant events and their possible alterations, and establishment of corresponding variables and assignments of their values to event-alterations – is missing in Pearl's account of model generation.

The decisive difference between Pearl's and Hitchcock's accounts of model generation is though the following one, which concerns the strength of the assumptions that Pearl and Hitchcock entertain, and with this also concerns the nature of their accounts:

Hitchcock implicitly stipulates that at the outset of the modeling procedure a complete counterfactual knowledge *K* about the situation *S* to be modeled is provided. Once the variables and the assignment of their values to event-alterations are established, Hitchcock's basic set of exhaustive counterfactuals is completely determined by this complete counterfactual knowledge *K*. Since the structural equations are extracted from this basic set of exhaustive counterfactuals, the structural equations also inherit this unequivocal determination. The important fact is that this procedure uniquely determines the structural equations for all admissible value distributions of the variables, in particular the non-actual ones. Summarizing, for Hitchcock there is exactly one causal model *M* that is unequivocally derived from a complete counterfactual knowledge *K* about the situation *S*, given the initial choice of variables and their value assignments. The only under-determination that remains in Hitchcock's account lies in the initial choice of event-alterations considered to be relevant in the situation *S*.

In this way, Hitchcock entertains the strongest assumption possible in a model generating procedure – he in essence stipulates the causal model *M*. In Hitchcock's own view, this may have seemed sensible, because he wanted to give a metaphysical account of causation anyway by reducing causation to certain kinds of counterfactual dependencies.[123]

However, in my view it is a severe disadvantage to stipulate a causal model in an account of causation. First, this does no longer allow to split the problem of analyzing causation in two sub-problems – the generation of a suitable causal model on the one hand, and the definition of causation qualified to such a model on the other. And as I have argued it is exactly this division of causal analysis into two sub-problems that marks a decisive progress of causal modeling accounts in comparison to classic metaphysical accounts of causation.

---

[123] Cf. in particular p. 274 and p. 287 of [cH 01].

Second, clearly Hitchcock's stipulation of a complete counterfactual knowledge K about a situation S to be modeled leaves two questions open: on the one hand, how such a stipulation can be justified, and on the other hand where this knowledge K comes from.
From these observations I draw the following conclusion regarding the first problem of causal modeling, as I describe it, namely the generation of a suitable causal model to formalize a given situation S: In my view, Hitchcock fails in his attempt to simplify Pearl's account of model generation. What Hitchcock provides is instead an account of model stipulation that just rewrites exactly the same counterfactual information that has been put in at the very beginning.

In striking contrast, Pearl's strategy for generating causal models is strictly pragmatic. Pearl is much more parsimonious in the assumptions that start his model generating procedure. He only assumes a set of random variables $\boldsymbol{U}\cup\boldsymbol{V}$ and some statistical data in the form of a joint probability distribution P over these variables to be given.
In the case that we are interested in analyzing causation on the type level and these variables $\boldsymbol{U}\cup\boldsymbol{V}$ formalize types of events and their possible alterations, such an assumption seems to be very reasonable. However, if we deal with causation on the token level and these variables $\boldsymbol{U}\cup\boldsymbol{V}$ formalize singular events and their possible alterations, the assumption of the existence of a joint probability distribution P over the set of variables $\boldsymbol{U}\cup\boldsymbol{V}$ is more problematic.
At least the following questions emerge: First, how can we make sense of a probability $p(u_1,\ldots u_n,v_1,\ldots,v_m)$ that is assigned to a combination of values $(u_1,\ldots u_n,v_1,\ldots,v_m)$ of the variables in $\boldsymbol{U}\cup\boldsymbol{V}$, when $(u_1,\ldots u_n,v_1,\ldots,v_m)$ formalizes a corresponding combination of possible alterations of *singular* events? Second, how can we observe such probabilities, if we are dealing with singular events, which we usually only observe once, if at all?

Certainly, Pearl would have to address these problems in order to argue that his account of model generation is applicable in a derivation of models for the analysis of token causation. And over and above this, also for the generation of models for the analysis of causation on the type level, Pearl would at least have to motivate where these probabilities come from and in which way they are observable. However, I think in the present context an investigation into the foundation of probability theory, in order to find out whether Pearl is justified in stipulating such a probability distribution P or not, would go too far.
Anyhow, this thesis is mainly concerned with the second discipline of causal modeling, as I call it, the definition of causation relativized to a given model. For this reason, I have only provided a brief synopsis of Pearl's account of model generation here, abstain from any criticism, and assume for the sake of Pearl's argumentation that his stipulation of a probability distribution P might be justified.

We have seen on the last pages that Pearl's general strategy for constructing a causal model starts out with a probability distribution P conveying conditional probabilistic dependencies, out of which causal patterns are extracted, and ends up with a multitude of basic causal structures D that all consistently formalize the given situation S at hand. Certainly, Pearl also assumes a sense of completeness of counterfactual information about this situation S to be modeled. Namely, a certain causal model M is fixed before causal analysis can take place, since the latter requires the construction of causal beams M_u out of this model M. And clearly, such a deterministic model M contains the complete counterfactual information about all the variables it contains, since for every possible value of the exogenous variables U_k, in particular non-actual values, the functions f_i uniquely determine the value of every endogenous variable V_i.

However, one can understand the basic causal structures D as blueprints for a complete causal model M, since these basic causal structures D certainly give criteria that a complete causal model M has to fulfill.

Summarizing, Pearl delivers a pragmatic account of model generation that employs probabilistic information about the given situation S to be modeled and does not stipulate the complete causal model as Hitchcock does. In this way, Pearl preserves the distinction of causal modeling into two disciplines – generating a suitable causal model for model and defining causation according to such a model.

A byproduct of this distinction of causal disciplines is again the following result that I already drew earlier: Pearl's definitions of actual and contributory causation, (P AC) and (P CC) respectively, only describe causation according to a certain previously chosen causal model M, and naturally refrain from any bold metaphysical claims about causation as such, as Hitchcock makes them.[124] Remarkably, Pearl himself seems not to be aware of this distinctive property of his account; for in his formulations (P AC) and (PCC), he calls an event a cause of another one, unqualified to a certain model M.[125]

[124] Cf. Hitchcock's definition of causation 'as such' on p. 287 of [cH 01].
[125] Cf. [jP 00], p. 319.

CHAPTER 3

PARADIGMATIC CASES JUDGED DIFFERENTLY

In the last two chapters, I have tried to clarify Hitchcock's account of actual and contributory causation, shortly discussed Woodward's notion of inclusive token causation, and described the relevant parts of Pearl's account of actual and contributory causation. In dealing with these accounts of Hitchcock, Woodward and Pearl, I examined the solutions that these accounts offer to the two basic problems of causal modeling. The first of these problems was the construction of an appropriate causal model M to describe a certain situation S containing singular events in whose causal relations we are interested. The second problem was to formalize the notion of token causation according to such a causal model, once we have agreed on a suitable one.

I maintained that it is a great advantage to generally relativize statements about token causation to certain models, since in this way these two problems can be tackled separately. Moreover, I motivated that Hitchcock's, Woodward's and Pearl's definitions of actual and contributory causation, or inclusive token causation respectively, do capture exactly such notions of causation that are relativized to a model.

However, since Hitchcock, Woodward and Pearl all, at least implicitly, claim to account for unqualified notions of causation, I briefly discussed Hitchcock's and Pearl's procedures of constructing appropriate causal models for analyzing token level causal dependencies. I reached the result that Hitchcock by assuming total information about the situation to be modeled tried to avoid the problem of model generation completely by in essence just stipulating a certain model. In contrast to this, Pearl did not assume total information about the situation and offered certain criteria that a suitable model had to fulfill. However these criteria of Pearl's were insufficient to determine a model on which his definition of actual and contributory causation could be applied.

In this chapter I only consider the second problem of causal modeling, formalizing the notion of token causation according to a causal model, once we have agreed on a suitable one. Concretely, I construct four paradigmatic examples to which I apply my reconstructions of Hitchcock's definitions of actual and contributory causation, (H AC) and (H CC) respectively, Pearl's corresponding definitions (P AC) and (PCC), and finally Woodward's definition of token causation (W TC).

This application of these different accounts to my four examples serves various purposes: First, it is easy to compare the extensions of the different accounts, i.e. one can easily determine whether a cause according to one definition is also a cause according to another one. Second, the verdicts of these accounts can be compared not only to each other, but can also be compared with our causal intuitions in these examples; this allows us to find out whether some definitions are more plausible than other ones. Third, possible formal difficulties in the application of these definitions can become obvious.

The focus of this application is on Hitchcock's account of actual and contributory causation with definitions (H AC) and (H CC), and on Pearl's account with the corresponding definitions (P AC) and (PCC). Again, I only marginally deal with Woodward's account, since his definition of token causation (W TC) is equivalent to Hitchcock's definition of contributory causation (H CC), and since Hitchcock's account was the earlier one and moreover in need of clarification. For the sake of simplicity, I will in the following treat my reconstructions of Hitchcock's definitions (H AC) and (H CC) as if they were Hitchcock's own version. My justification for this treatment is that I consider my reconstruction of Hitchcock's account in the first chapter of this thesis to be the straightforward formal expression of his rationale, moreover staying as closely as possible to the original.

This chapter has two leading questions: The first question is whether Hitchcock, and in this way Woodward, too, succeed in giving a simplification of Pearl's account of singular event causation. The second question is whether one can improve on their accounts. In the subsequent eight sections I answer these questions as follows:

In the first section I show that Hitchcock's project of simplifying Pearl's account of actual causation fails. I offer two examples. Example 1 contains an actual cause according to Hitchcock's (H AC) that is not a cause according to Pearl's (P AC). Example 2 contains an actual cause according to (P AC) that does not qualify as a cause for (H AC). In both examples the verdict delivered by Pearl's account is in accord with our causal intuition, whereas Hitchcock's account reaches verdicts that contradict our intuitions. Apart from that, I make the observation that Pearl's account of actual causation is not directly applicable to these examples for technical reasons, since it does not allow interventions on exogenous variables.

The second section deals with a revocation of his account of token causation that Hitchcock makes in a footnote in a later article. This article implicitly features two new notions of component and net causation. I offer formalizations of token level reductions of these notions. Moreover, I argue that my token level reductions of these new notions cannot replace the basic notion of actual causation, since the latter is deeply entrenched in our causal intuition. Hence, I unify my formalizations of token component causation and token net causation to a definition of merged actual causation (H MAC). An application of this new definition to my two examples from the first section, though, shows that (H MAC) still fails in one of them.

In the third section I offer a natural extension of Pearl's account of actual and of contributory causation that allows interventions on exogenous variables. I achieve this by slight modifications in Pearl's definitions of a causal beam (P CB), a natural beam (P NB), and of actual and contributory causation (P AC) and (P CC). These modifications facilitate a uniform treatment of exogenous and endogenous variables that is in my view more natural than Pearl's original distinction between exogenous and endogenous variables in constructing sustaining sets.

In the fourth section I construct canonical extensions, called example 3 and 4 respectively, of my two examples featured in the first section. This is achieved by introducing a parallel mechanism that symmetrically over-determines the putative effect in examples 1 and 2. Subsequently, I apply Hitchcock's account of contributory causation given by definition (H CC) and my extension of Pearl's account of contributory causation given by my definition (EP CC) to examples 3 and 4. Hitchcock's definition (H CC) analyzes example 3 as containing a contributory cause that is not a cause according to my extension of Pearl (EP AC). In example 4 the analyses yielded by both definitions (H CC) and (EP AC) coincide in the example containing a contributory cause. The verdict delivered by my account is in both examples in accord with what I take to be our causal intuition, whereas Hitchcock's account reaches a verdict in example 3 that contradicts our intuitions again.

The fifth section contains an examination of the reason why Hitchcock's account of actual and contributory causation fails in the analysis of my examples. I conclude that this reason can be found in Hitchcock's underlying rationale that identifies a causal process linking a cause and its effect with a route in a causal model. I argue that this understanding of a process is misguided, since the notion of a causal process linking a cause with its effect in a situation requires a structurally more complicated construction to be formalized. I maintain that such a construction is given by Pearl's definition of a causal beam that also captures a notion of sustenance.

In the sixth section I apply Woodward's account of inclusive token causation given by (W TC) to my four examples. As observed in Chapter 1, Woodward understands (W TC) to express an inclusive notion that uniformly applies to all cases of singular event causation. I show that Woodward's inclusive understanding of token causation is an improvement over Hitchcock's separated notions of actual and contributory causation, since in example 2 Woodward's account with (W TC) delivers a result that is in accord with our causal intuition; Hitchcock's account of actual causation with (H AC) and (H AR) previously failed with this example. However, Woodward's account with (W TC) still analyzes my examples 1 and 3 in a way yielding results that contradict our causal intuition.

In the seventh section I investigate the reason behind this failure of Woodward's account of inclusive token causation. I argue that this is still the same as in Hitchcock's case, since Woodward's definition (W TC), despite being applicable to all cases of singular event causation, is still equivalent to Hitchcock's definition of contributory causation (H CC). This failure of Woodward's account of token causation and the fact that it fails for the same reason as Hitchcock's account of actual and contributory causation finally justifies the fact that I discuss Woodward's account only subordinately in this thesis. Furthermore, I give a trivial extension of Pearl's account that captures an inclusive notion of token causation as well, in this way also covering all cases of singular event causation. In my view, the distinction between an inclusive notion of token causation and exclusive notions of actual and contributory causation is, though, a mere terminological problem.

The eighth section is a final proving ground for my extension of Pearl's account of actual and contributory causation given by my definitions (EP AC) and (EP CC). I demonstrate that my extension does not fare worse than Hitchcock's original account in dealing with the three major problems of Lewis's classic counterfactual account: the intransitivity of causation, preemption, and symmetric over-determination. For this reason, I apply my extension of Pearl's account to the examples 'Boulder' and 'Backup' that Hitchcock featured to illustrate intransitivity and preemption respectively, and finally to Lewis's classic example 'Firing Squad' to give a further worthy representative for cases of symmetric over-determination.

ACTUAL CAUSATION: HITCHCOCK'S ACCOUNT AND PEARL'S ACCOUNT APPLIED TO TWO EXAMPLES

In this section I examine whether Hitchcock's project of simplifying Pearl's account of actual causation succeeds. I reach the conclusion that Hitchcock's account fails in two respects: first by having a different extension than Pearl's account and second by contradicting our pre-theoretic causal intuitions.

I present two examples: The first contains an actual cause according to Hitchcock's definition (H AC) that is not an actual cause according to Pearl's definition (P AC). The second example contains an actual cause according to Pearl that does not qualify as an actual cause for Hitchcock. In both examples the verdict delivered by Pearl's account is in accord with what I take to be our causal intuition. Hitchcock's account reaches verdicts that contradict our intuitions.

Remarkably, it becomes obvious that Pearl's account of actual causation is not directly applicable to these examples for technical reasons, since Pearl's definitions of a natural beam (P NB) and of actual causation (P AC) do not allow interventions on exogenous variables. I overcome this obstacle by the introduction of dummy variables.

The first example demonstrates that actual causation according to Hitchcock does not imply actual causation according to Pearl by containing an actual cause according to (H AC) that is not an actual cause according to (P AC).

Example 1

Cover story (Atomic Binding): An experiment is carried out in which the binding process of protons and electrons is examined. It is conducted in the following way: when a proton is released in a void and neutrally charged test chamber, a device releasing an electron in the test chamber is triggered, so that the following holds: Whenever a proton is released, an electron is released; when there is no proton, there is also no electron. A perfect detector records the emission of the particles in the test chamber (also of the hydrogen atom that spontaneously forms); in particular, it records the resulting electric charge of the test chamber

Assume that we are particularly interested in the question whether the release of the proton affects the electric charge of the system. Such a question is not as nonsensical as it might appear on first glance. In a calibrating scenario, when we want to compare the behavior of the detector in a certain test series to its ideal behavior, questions like this are the decisive ones.

The following assignment of events to values of variables

$$P = \begin{cases} 0, & \text{if Proton is not released into the test chamber} \\ 1, & \text{if Proton is released into the test chamber} \end{cases}$$

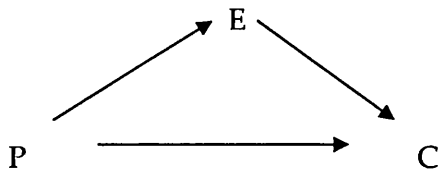
$$E = \begin{cases} 0, & \text{if Electron is not released into test chamber} \\ 1, & \text{if Electron is released into test chamber} \end{cases}$$

$$C = \begin{cases} 0, & \text{if according to detector electric Charge of system does not stay constant} \\ 1, & \text{if according to detector electric Charge of system does stay constant} \end{cases}$$

yields the functions, or structural equations in Hitchcock's terminology,

$$f_E(p) = id_P \quad \text{and} \quad f_C(e,p) = \begin{cases} 0, & \text{if E and P have different values} \\ 1, & \text{if E and P have the same value} \end{cases}$$

with the associated graphical representation:

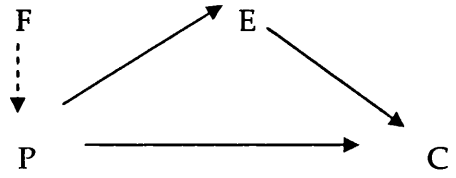


Let us assume that the actual value of P is $P = 1$, so that the resulting values are $E = 1$ and $C = 1$.

An application of Hitchcock's account proceeds in the following way: Demonstrate that route (P,C) is active. Intervene on variable P, setting $P = 0$ and freeze variable E at its actual value $E = 1$. This yields $C = 0$ showing that the direct route (P, C) is active and that $P = 0$ is an actual cause of $C = 0$.

In order to apply Pearl's account meaningfully and non-trivially to this example, one first has to add a dummy parent variable F of P. For, as remarked in the last chapter, values of exogenous variables can never qualify as causes for Pearl's account. So, add the variable F and a new function for variable P,

$f_P(f) = id_F$, yielding the following diagram:



Heuristically, variable F could be understood to be modeling the possible states of the device generating and emitting the protons, i.e.

$$F = \begin{cases} 0, & \text{if the proton releaser is not triggered} \\ 1, & \text{if the proton releaser is triggered} \end{cases}$$

For the present solution in the model with $P = 1$, $E = 1$ and $C = 1$, corresponding to model state $F = 1$, there is only the trivial natural beam that is identical to the original model: For, according to Pearl's definition of a causal beam (P CB), the only possible choice of sustaining sets for the variables E and C are

$$\mathbf{S}_E = \{P\},\ \underline{\mathbf{S}}_E = \varnothing \quad \text{and} \quad \mathbf{S}_C = \{P, E\},\ \underline{\mathbf{S}}_C = \varnothing.$$

That means no variables are frozen, so that the resulting projection functions of the beam are still the original functions of the causal model, i.e. $f_E^{F=1}(p) = f_E(p)$ and $f_C^{F=1}(p,e) = f_C(p,e)$. Obviously, these projection functions are non-trivial in their arguments, as required in definition (P CB). In this way, the solution in the beam $M_{F=1}$ for $P = 1$ is $C = 1$, and an intervention setting $P = 0$ in this beam $M_{F=1}$ still yields $C = 1$. According to Pearl's definition of actual causation (P AC), $P = 1$ is hence not an actual cause for $C = 1$.

As a result, Pearl's account accords with our pre-theoretic intuition in this example, or at least mimics it, whereas the verdict that Hitchcock's analysis of this example delivers contradicts this intuition. For, the detector reading stays constant, no matter whether there is a proton released into the test chamber or not. In this way, it is not only shown that actual causation according to Hitchcock does not imply actual causation according to Pearl. The example also provides strong evidence against the correctness of Hitchcock's account suggesting that it is too wide in certain respects, namely in characterizing events as causes that are not causes according to our natural understanding.

The next example demonstrates that the inverse implication from causation according to Pearl to causation according to Hitchcock is not valid either. It contains an actual cause according to Pearl's account that is not an actual cause under Hitchcock's analysis.

Example 2

Cover story (radioactive decay): A radioactive atom *a* is placed in an experimental set-up with a perfect detector. When atom *a* decays into two smaller atoms *s* and *w*, the detector registers both *s* and *w*. The detector is furthermore equipped with an analysis program that gives the result 'decay' if the detector has been triggered by two atoms.

Suppose that we are interested in the question whether the decay or non-decay of the atom has an effect on the detector. In particular, the focus is on the question whether a non-decay of the atom causes the detector not to give the result 'decay'. Again, this is a typical question for calibrating the detector, i.e. when we want to adjust the real behavior of the detector in certain test series to its ideal behavior represented in the model.

Assignment events to values of variables as follows:

$$T = \begin{cases} 1, & \text{if atom a decays in particles s and w} \\ 0, & \text{otherwise} \end{cases}$$

$$S = \begin{cases} 1, & \text{if particle s is registered by detector} \\ 0, & \text{otherwise} \end{cases}$$

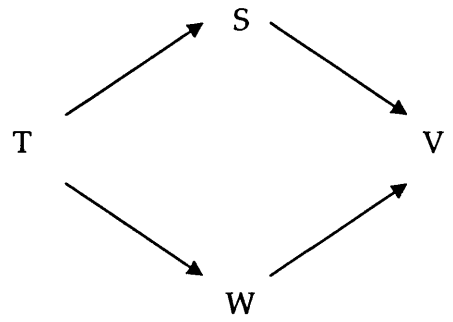
$$W = \begin{cases} 1, & \text{if particle w is registered by detector} \\ 0, & \text{otherwise} \end{cases}$$

$$V = \begin{cases} 1, & \text{if detector says 'decay'} \\ 0, & \text{otherwise} \end{cases}$$

This yields the functions $f_S(t) = id_T$, $f_W(t) = id_T$, and

$$f_V(s,w) = \begin{cases} 1, & \text{if } S=1 \text{ and } W=1 \\ 0, & \text{otherwise} \end{cases}.$$

With this, the associated graphical representation of the model is:

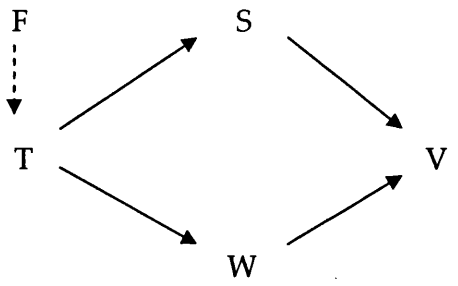


Assume that the actual state of the model is $T = 0$ with the resulting solution $S = 0$, $W = 0$ and $V = 0$.

An application of Hitchcock's account yields the result that neither the route (T, S, V) nor the route (T, W, V) is active: Intervening on T setting $T = 1$ while freezing W at its actual value $W = 0$ does not alter the value of V, rendering route (T, S, V) inactive. Analogously, setting $T = 1$ while freezing S at $S = 0$ does not alter the value of V, rendering route (T, W, V) inactive, too. As a result, $T = 0$ is not an actual cause of $V = 0$ according to Hitchcock.

Again, a meaningful application of Pearl's account requires the addition of a dummy parent variable F of T. So, add again the variable F and a new function for variable T,

$F_T(p) = id_F$ yielding the diagram:



According to Pearl, $T = 0$ is a cause of $V = 0$, if and only if there is a natural beam in which an intervention on T setting $T = 1$ brings about $V = 1$. Such a natural beam is given by the trivial beam that is identical to the original model: An admissible choice of sustaining sets for the variables S, W and V, obviously fulfilling clauses (i) and (ii) of Pearl's definition (P CB), is

$$\mathbf{S}_S = \{T\},\ \underline{\mathbf{S}}_S = \varnothing; \quad \mathbf{S}_W = \{T\},\ \underline{\mathbf{S}}_W = \varnothing; \quad \text{and} \quad \mathbf{S}_V = \{S, W\},\ \underline{\mathbf{S}}_V = \varnothing.$$

Again, no variables are frozen, so that the resulting projection functions of the beam $M_{F=1}$ are still the original functions of the causal model, i.e. $f_S^{F=1}(t) = f_S(t)$, $f_W^{F=1}(t) = f_W(t)$ and $f_V^{F=1}(s,w) = f_V(s,w)$. Again, the projection functions $f_S^{F=1}$, $f_W^{F=1}$ and $f_V^{F=1}$ are clearly non-trivial in their respective arguments according to definition (P CB), since the original functions f_S, f_W and f_V in model M were non-trivial. In this beam $M_{F=1}$ intervening on T setting $T = 1$ brings about $V = 1$, so that according to Pearl's definition (P AC), $T = 0$ is an actual cause of $V = 0$.

Again, our pre-theoretic intuition accords with Pearl's account and speaks against Hitchcock's analysis of this example, for clearly the net effect of the decay of atom *a* is the detector reading 'decay', and also in the complementary case, a non-decay of the atom is responsible for the detector not giving the reading 'decay'. Thus also the inversion of the above non-implication is shown: causation according to Pearl does not imply causation according to Hitchcock. And again, strong evidence is provided against the correctness of Hitchcock's account, this time suggesting that it is also too narrow in certain respects, namely in not characterizing events as causes that are causes in our natural understanding.

Summarizing, Hitchcock's project of simplifying Pearl's account of actual causation fails in two respects: First, Hitchcock's account has a different extension than Pearl's. In itself, this would not necessarily be a bad thing, if the differences in judgment of these two accounts were in cases in which both verdicts were plausible. However, they are not, and this is the second and grave shortcoming of Hitchcock's project. In my chosen examples our causal intuitions are more or less unambiguous, and the verdicts delivered by Hitchcock's account contradict them. What makes my examples even more interesting is that they are structurally very basic, such that this violation of our causal intuition already takes place on a fundamental level. In this understanding, example 2 illustrates that Hitchcock's account already fails when it has to analyze certain basic net effects.

HITCHCOCK'S LATER NOTIONS OF NET AND COMPONENT CAUSATION

Admittedly, Hitchcock retracted his account featured in this thesis in a footnote in a later article, called 'A Tale of Two Effects'.[126] In the same footnote, he mentions that his account featured in this thesis should better be understood as capturing a notion of actual component causation.

In 'A Tale of Two Effects' Hitchcock introduces this notion of component causation, or causation along a route, and contrasts it with a notion of net causation. Hitchcock himself only talks of component effects and net effects and prefers not to talk of the corresponding division of actual causation into component and net causation; in my view, though, such a division into two kinds of actual causation is canonically induced by a division into two kinds of effects, when these effects are token effects. Hitchcock motivates his new distinction at length, but does not give a formal definition of these new concepts.

This can easily be done, though, and in the following I offer a formal reconstruction of these two concepts. Admittedly, this reconstruction is only a crude one. Formally, it is kept close to Hitchcock's original account mainly featured in this thesis, in order not to deviate unnecessarily from the basic concepts introduced by Hitchcock. I also made no attempt to cover Hitchcock's newly won intuitions about component and net causation. The main purpose of my formal reconstruction is to deduce judgments about my two examples from Hitchcock's new account.

The straightforward definition of actual net causation is in my view the following:

---

[126] Cf. [cH 01b], p. 363.

Definition (Actual Net Causation, Hitchcock) **(H ANC)**

Let c and e be distinct occurring events in a situation S and let $M := \langle V, E\rangle$ be a causal model formalizing S, such that the following holds: There are variables X and Z in V, such that the values x of X and z of Z in the actual solution $R(V)$ of $\langle V, E\rangle$ represent the actually occurring events c and e in S and such that their non-actual values x' and z' represent alterations of c and e respectively.

Then c is an *actual net cause* of e according to the model M if and only if in the actual solution $R(V)$ of M the value z of Z counterfactually depends on the value x of X.

In this context, Hitchcock's original definition of counterfactual dependence (H CDSCM) is taken over without any changes. I.e. still the value z of variable Z depends counterfactually on the value x of variable X in the solution $R(V)$ of the model $M = \langle V, E\rangle$ if and only if the following holds: In the solution $R(V)$ of $\langle V, E\rangle$ it is $X=x$ and $Z=z$, and there are values $x'\neq x$ and $z'\neq z$, such that replacing the equation Ex for X with the new equation $E'x := X=x'$ in E yields the result $Z=z'$ for variable Z.

As a trivial result, Hitchcock's new actual net causation reduces to simple counterfactual dependence. No variables are frozen anymore in the model while intervening on variable X. The change in the value of X brought about by the intervention propagates through the unaltered model to variable Z.

Clearly this is a strong deviation from Hitchcock's previous idea to isolate causal routes r between variables X and Z in a causal model by freezing the variables that are intermediate between X and Z but do not belong to route r. Hitchcock's original rationale that c causes e if and only if these events are linked by a process, equating such a process to a route between the associated variables X and Z in an appropriate model, has hence to be

given up. If one still wants to keep this intuition of a process linking cause and effect, the only straightforward candidate for a formal expression of such a process would hence be the causal model itself. But as said before, this is not the right place for examining these issues more deeply.
Interestingly, though, Hitchcock's understanding of causal processes as routes in a causal model is preserved in his new notion of actual component causation. A formalization of this new notion differs only in one respect from Hitchcock's definition of causation (H AC) featured in this thesis, namely by letting a set of routes be causally active instead of a single route:

Definition (Actual Component Causation, Hitchcock) **(H ACC)**

Let c and e be distinct occurring events in a situation S, and let $M := \langle V, E \rangle$ be a causal model formalizing S, such that the following holds: There are variables X and Z in V, such that the values x of X and z of Z in the actual solution $R(V)$ of $\langle V, E \rangle$ represent the actually occurring events c and e in S and such that their non-actual values x' and z' represent alterations of c and e respectively.
Then c is an *actual component cause* of e according to the model M if and only if in the actual solution $R(V)$ of M there is an active set of causal routes from X to Z.

In this context, Hitchcock's definition (H AR) of an active route is canonically extended to apply to a set of routes:

Definition (Active Set of Routes, Hitchcock) **(H ASR)**

A *set of routes* $\{(X,Y_1,\ldots,Y_n,Z), (X,V_1,\ldots,V_n,Z), (X,W_1,\ldots,W_n,Z), \ldots\}$ is *active* in a solution $\boldsymbol{R}(\boldsymbol{V})$ of the causal model $\langle \boldsymbol{V},\boldsymbol{E}\rangle$ if and only if the actual value z of Z depends counterfactually upon the actual value x of X within the resulting solution $\boldsymbol{R}'(\boldsymbol{V})$ of the new system of equations $\boldsymbol{E}'$ constructed from $\boldsymbol{E}$ as follows:
For all Y in $\boldsymbol{V}$, if Y is intermediate between X and Z but does not belong to the set of routes $\{(X,Y_1,\ldots,Y_n,Z), (X,V_1,\ldots,V_n,Z), (X,W_1,\ldots,W_n,Z), \ldots\}$, then replace the equation E_Y for Y with a new equation $E'_Y := Y=y$ that sets Y constant to its value y in the solution $\boldsymbol{R}(\boldsymbol{V})$.

Again, counterfactual dependence is defined by the original (H CDSCM); and again, if there are no intermediate variables that do not belong to this set of routes, then $\boldsymbol{E}'$ is just $\boldsymbol{E}$.
For reasons of space I do not discuss Hitchcock's motivation for this new distinction between actual net and actual component causation, nor do I make any attempt to reconstruct the details of this later view of his. The main purpose for me to discuss this changed view of his is to examine how this new distinction fares with respect my two examples above. More precisely, I settle for a formal reconstruction of what it means that c is an actual net cause, or actual component cause, of e according to a given model. Certainly, I kept the notion of an appropriate model in my definitions (H ANC) and (H ACC) above, but only to stay as close as possible to Hitchcock's original formulation in (H AC). Nevertheless, I am only concerned here with the second task of causal modeling as I understand it: the definition of causation according to a model. That means I deliberately neglect whether Hitchcock's view regarding the first task of causal modeling in my understanding, the problem of generating an appropriate causal model, has changed in his later account. My

justification for this neglect is that I could not find any evidence in Hitchcock's article that this has changed.

Returning to the problem for the original Hitchcock account posed by my two examples above, the first immediate observation to be made is that Hitchcock's new account simply tries to avoid this problem by not talking about actual causation as a generic concept any more. Instead, Hitchcock tries to motivate that there are only different aspects of causation, not a uniform generic understanding of one single causal relationship. Applied to our modeling problem on the token level, this would imply that talk of one concept of actual causation according to a model is not meaningful.
I do not want to enter into a discussion of this topic any deeper. For my present concern, it suffices to state that I hold the firm conviction that we have a clear pre-theoretic intuition about actual causation; and this intuition usually delivers unequivocal judgments about situations S when we ask whether a certain event c in S actually causes another event e in S. Moreover, I regard the rationale underlying Pearl's and the original Hitchcock account as very plausible and convincing: We say that a certain event c in S actually causes another event e in S if and only if there is a causal process, i.e. a concatenation of the mechanisms or local laws at work in situation S, linking c and e such that c sustains e.
If we insist in this way on the generic notion of an actual cause according to a model, there is only one straightforward way to explicate this notion in Hitchcock's later view, namely to merge actual net causation and actual component causation:[127]

[127] Certainly, Hitchcock would not agree to such a formulation, since he has to maintain that a generic notion of actual causation cannot be constructed out of two disjoint aspects of net causation and component causation.

Definition (Merged Actual Causation, Hitchcock) **(H MAC)**

Let c and e be distinct occurring events in a situation S, and let $M := \langle V, E \rangle$ be a causal model formalizing S, such that the following holds: There are variables X and Z in V, such that the values x of X and z of Z in the actual solution $R(V)$ of $\langle V, E \rangle$ represent the actually occurring events c and e in S and such that their non-actual values x' and z' represent alterations of c and e respectively.

Then c is a *merged actual cause* of e according to the model M if and only if the following holds:

c is an actual net cause of e according to M, or

c is an actual component cause of e according to M.

The analysis of example 2 by the definition of merged actual causation (H MAC) now delivers a verdict that no longer contradicts our pre-theoretic intuition about actual causation in this example:

$T = 0$ is identified as a merged actual cause of $V = 0$ according to (H MAC). For, an intervention in model $\langle V, E \rangle$ on variable T setting $T = 1$ yields $S = 1$ and $W = 1$, and in turn $V = 1$, demonstrating that $V = 0$ is counterfactually dependent on $T = 0$. With this, V is counterfactually dependent on T in the actual set of structural equations E in Hitchcock's terminology, such that according to my definition (H ANC) $T = 0$ is an actual net cause of $V = 0$. With (H MAC), $T = 0$ is lastly a merged actual cause of $V = 0$.

And this result is in accord with our intuition about actual causation in this case, since we consider $V = 0$ to be a clear net effect of $T = 0$.

However, the analysis of example 1 with the artificially defined notion of merged actual causation in (H MAC) still delivers a verdict that contradicts our pre-theoretic intuition about actual causation in this example:

$P = 0$ is identified as a merged actual cause of $C = 0$ by (H MAC). For, the originally active route (P,C) trivially constitutes a singleton set {(P,C)} of active routes. More elaborately, an intervention on variable P, setting $P = 0$ and freezing variable E at its actual value $E = 1$ yields $C = 0$. With this, in Hitchcock's terminology the value of variable C is counterfactually dependent on the value of variable P in the new set of structural equations E' achieved by freezing variable E at $E = 1$. As a result, the singleton set {(P,C)} is active, and according to my definition (H ACC) $P = 0$ is an actual component cause of $C = 0$. Again with (H MAC), $P = 0$ is finally a merged actual cause of $C = 0$.
And as mentioned above, our intuition about actual causation in this case rules that $P = 0$ is not responsible for $C = 0$, since the resulting value of C is $C = 0$, no matter whether $P = 0$ or not.

Summarizing, my notion of merged actual causation expressed by (H MAC), which I artificially constructed out of token level reductions of Hitchcock's two aspects of causation, still does not accord with our pre-theoretic intuitions about actual causation. Admittedly, the construction of this notion by (H MAC) is a little farfetched. However, the question remains whether one wants to give up the talk of actual causation as a generic term and settle for a division in component causation, net causation and possibly further notions. So far, I see no need to give up a generic notion of actual causation according to a model. For, as already mentioned, we have a clear pre-theoretic intuition about this in my view, and the understanding of actual causation as a process sustaining an effect in a situation brings further credibility to this intuition. And as I have demonstrated above, Pearl's account formalizing this idea of an actual cause being a sustaining process can handle my examples. From these considerations, I see only reasons against renouncing actual causation according to a model.

To put these notions of actual component causation and actual net causation of Hitchcock's into perspective, consider them from the viewpoint of the distinction of actual and contributory causation. I have said before that contributory causation is a version of actual causation that applies to cases in which a token effect is symmetrically over-determined by various token causes. This was my attempt to suggestively express the fact that both concepts have the same underlying rationale in my view and that contributory causation is the more fundamental concept of the two. For, following the understanding of a token cause as being a causal process sustaining its token effect against possible contingencies, it became clear that the constraints posed by these contingencies for actual causes are much more specific and harder to meet than they are for contributory causes. Be this as it may, my main point here is that these two concepts of contributory and actual causation share their basic rationale.

In strong contrast to this, actual component causation and actual net causation do not share an underlying idea: Actual component causation tries to isolate causal processes in a situation by focusing on routes in the causal model, actual net causation relies on simple counterfactual dependence. With this, actual component causation and actual net causation simply express other aspects of causation that may very well be justified and in which we may very well be interested, too. However, as seen above, they cannot replace the notion of actual causation. And, as I have said this many times, we not only have a firm intuition about actual causation but also a strong interest in it.

AN EXTENSION OF PEARL'S ACCOUNT OF ACTUAL AND CONTRIBUTORY CAUSATION

Let me return now to the way my examples 1 and 2 are analyzed by Pearl's account of actual causation. As became obvious, in both cases dummy variables had to be introduced, such that an intervention on variables that were previously exogenous was possible. Such an introduction of dummy variables is heuristically more problematic than it might seem on first glance, though. Consider the assignment of certain event-alterations to the values of dummy variable *F* in both examples: In example 2 we could interpret *F* as modeling a bombardment of atom *a* with a certain particle, thus inducing nuclear fission. However, then we would have a case of nuclear fission and no longer a proper case of spontaneous nuclear decay, that means we would have altered the nature of the situation to be modeled significantly. That is exactly what we implicitly achieved in our treatment of example 1, altering the situation by identifying the variable *F* with the device producing protons that was not mentioned in the cover story.

As a result, Pearl's account of actual causation faces the following dilemma: Either dummy variables are identified with event-alterations in an extended situation, or dummy variables are understood as simple formal devices without a practical meaning. Both options are more than undesirable.

The first option would yield the result that Pearl's account is simply not applicable to any cases in which we are interested in whether a value of an exogenous variable is cause of another value of another variable. For, introducing an additional variable in the model would clearly distort the original situation. The second option would leave the modeled situation intact, but is formally highly questionable. On the one hand, an arbitrary

introduction of variables that do not represent any event-alterations in the situation is completely unmotivated and ad hoc. On the other hand and much more gravely, the underlying rationale that a cause is a sustaining process between the cause and the effect variable is no longer compatible with constructing a beam according to a modified model in which a dummy variable has been introduced.

Since I see no plausible way for Pearl's account out of this dilemma, I propose a straightforward extension of Pearl's original account, which easily overcomes the problem of the inadmissibility of values of exogenous variables as causes. Starting point for this extension is the observation that for the construction of a causal beam according to Pearl's definition (P CB) the following holds:

The systematic addition of a dummy variable D_j to every exogenous variable U_j for j=1,...,m in a causal model M, also augmenting the model by the function $f_{U_j}(d_j)= id_{D_j}$ and then constructing the associated beams M'_d for this extended model M' and state d of D corresponds to the following procedure.[128] Consider the original model M in state u, but in constructing the associated beam M_u, admit the exogenous variables U in the sustaining sets $S \subseteq PA_i$.

Such a uniform treatment of exogenous and endogenous variables in constructing sustaining sets would in no way distort Pearl's original rationale of causal beams as projections on fixed states. The only difference is that the projection functions f^u in M_u would be responsive to changes in the values of U. This would allow interventions on exogenous variables in a causal beam to have an effect on the values of its endogenous variables, such that values of exogenous variables could be causes in the same way as values of endogenous variables can be.

[128] There, for every j=1,...,m, it is $d_j = id^{-1}{}_{D_j}(u_j)$, the argument of the function id_{D_j} for the value u_j, (hence not assuming that function id can automatically be inverted).

My extended version of Pearl's account of actual causation implements this uniform treatment of exogenous and endogenous variables in Pearl's original definitions of a causal model, of a causal beam and of actual causation. Only slight modifications are needed to achieve this aim. The explicit construction of this extended account is given as follows:

Definition (Causal Model, Extended Pearl) **(EP CM)**

A causal model is a triple $M = \langle \boldsymbol{U}, \boldsymbol{V}, \boldsymbol{F} \rangle$, where

(i) $\boldsymbol{U}$ is a set of background variables, also called 'exogenous', whose values are causally determined by factors outside the model;

(ii) $\boldsymbol{V}$ is a set $\{V_1,\ldots,V_n\}$ of variables called 'endogenous', whose values are determined by variables in the model, i.e. by variables in $\boldsymbol{U} \cup \boldsymbol{V}$; and

(iii) $\boldsymbol{F}$ is a set of functions $\{f_1,\ldots,f_n\}$, such that each f_i is a mapping from the respective value-domains of $\boldsymbol{U} \cup (\boldsymbol{V} \setminus V_i)$ to that of V_i and such that the entire set $\boldsymbol{F}$ forms a mapping from $\boldsymbol{U}$ to $\boldsymbol{V}$ and the entire set $\boldsymbol{F}$ has a unique solution. Symbolically, the set of equations $\boldsymbol{F}$ can be represented by writing

$$v_i = f_i(\boldsymbol{pa}_i), \quad i = 1,\ldots, n,$$

where $\boldsymbol{pa}_i$ is any realization of the unique minimal set of variables $\boldsymbol{PA}_i$ in $\boldsymbol{U} \cup (\boldsymbol{V} \setminus V_i)$ sufficient for representing f_i.

The only change here is the relaxation of clause (iii), now unifying exogenous and endogenous variables in one parent set $\boldsymbol{PA}_i$. The decisive modification takes place in the definition of an extended causal beam.

Definition (Causal Beam, Extended Pearl) **(EP CB)**

For causal model $M = \langle U, V, \{f_i\}_{i=1,\ldots,n}\rangle$ and state $U = u$, a causal beam is a new model $M_u = \langle u, V, \{f_i^u\}_{i=1,\ldots,n}\rangle$ in which the set of functions f_i^u is constructed from $\{f_i\}_{i=1,\ldots,n}$ as follows:

1. For each variable $V_i \in V$, partition PA_i into two subsets $PA_i = S_i \cup \underline{S}_i$, where S_i is any subset of PA_i satisfying

   $$f_i(S_i(u), \underline{s}_i) = f_i(S_i(u), \underline{s}_i')$$

   for all realizations $\underline{s}_i$ and $\underline{s}_i'$ of $\underline{S}_i$.
   In words, S_i is any subset of PA_i sufficient to entail the actual value of $V_i(u)$ in the original model M, regardless of how we set the other members of PA_i.

2. For each variable $V_i \in V$, find a subset W_i of $\underline{S}_i$, for which there exists some realization $W_i = w$ that renders the function $f_i(s_i, S_{iw}(u))$ nontrivial in s_i; that is

   $$f_i(s_i', \underline{S}_{iw}(u)) \neq f_i(s_i, \underline{S}_{iw}(u)) = V_i(u)$$

   for some realizations s_i' of S_i.
   Here, for a set $Y \subseteq V$ and a set $W \subseteq (U \cup V)$, $Y_w(u)$ denotes the solution for Y of the set of equations $\{f_i : \{V_i \notin W\} \cup \{W = w\}\}_{i=1,\ldots,n}$.

3. Replace $f_i(s_i, \underline{s}_i)$ by its projection $f_i^u(s_i)$, which is given by

   $$f_i^u(s_i) = f_i(s_i, \underline{S}_{iw}(u)).$$

   Here, $\underline{S}_i$ should be chosen not to intersect the sustaining set S_j of any other variable V_j, for $j \neq i$. Likewise, setting $W_i = w$ should not contradict any other setting W_j for $j \neq i$.

In this way, an extended causal beam is still a projection of the original model M on the state u. However in an extended causal beam, the new parent sets $PA_i^u = S_i$ of V_i can contain exogenous variables $U_{k1}, \ldots, U_{kp}$, so that the function f_i^u is responsive to possible changes of the values of $U_{k1}, \ldots, U_{kp}$ by an intervention.
The definitions of an extended natural beam and of an extended actual cause are analogous reformulations of Pearl's original definitions (P NB) and (P AC), respectively:

Definition (Natural Beam, Extended Pearl) **(EP NB)**

An extended causal beam M_u is said to be natural if condition 2. in the above definition (EP CB) is satisfied with $W_i = \varnothing$ for all $V_i \in V$.

Definition (Actual Cause, Extended Pearl) **(EP AC)**

For extended model M, event $X = x$ is said to be an actual cause of $Y = y$ in a state u if and only if there exists an extended natural beam M_u such that:

$$Y_x = y \text{ in } M_u, \qquad \text{(C1)}$$

and

$$Y_{x'} \neq y \text{ in } M_u \text{ for some } x' \neq x. \qquad \text{(C2)}$$

Now, variable X can very well be exogenous, thus allowing interventions on exogenous variables. This is achieved by the last extended definition of $Y_w(u)$ in (EP CB) – the solution for Y of the set of equations $\{f_i : \{V_i \notin W\} \cup \{W=w\}\}_{i=1,\ldots,n}$ – now also allowing exogenous variables in the set W.

Apparently, these definitions bring about exactly the desired property, namely treating exogenous variables in exactly the same way as endogenous variables, thus allowing the value of an exogenous variable X to be a cause in exactly the same way as if X were endogenous.

One interesting problem remains though: It is not clear whether this new extended account of actual causation is conservative in characterizing endogenous variables as causes. For, definitions (P CB) and (EP CB) will usually produce structurally different beams to be evaluated by definitions (P AC) and (EP AC), respectively. This question is, though, not pursued further here. In the present context, it suffices to note that Pearl's underlying idea of sustenance of a value of a variable in a projection of a model to its states is preserved and that my extension is at least as natural as Pearl's original account: first, by treating all variables uniformly, no matter whether exogenous or endogenous, and second and more importantly, by being able to account for examples for which we clearly have a natural causal intuition.

It is only my conjectured rationale for Pearl's original account not allowing interventions on exogenous variables, since their determining mechanisms are not known, that would have to be given up, rendering interventions independent of our knowledge. Such an abstract understanding of interventions as hypothetical operations without any meaning outside the formal framework is, though, not implausible. After all, as I noted before, the interventions in Pearl's account are also not achieved by the corresponding mechanisms, i.e. the functions in the model. On the contrary, the function f_X of the variable X intervened on is overridden by the intervention.

Moreover, the impression arises that Pearl himself did not allow interventions on exogenous variables mainly for the sake of a simplified formal notation and a prevention of counter-intuitive results from the ambiguities in dealing with exogenous variables in the construction of sustaining sets. However, my relaxation of (P CB) to (EP CB) allows constructing sustaining sets completely unequivocally when Pearl's original definition (P CM) is relaxed to my (EP CM).
Overall, in my view parent variables can very well be exogenous according to our pre-theoretic intuition, and it seems to be much more natural to allow interventions on exogenous variables, too, thus treating all variables in the same way.

After this extended version of Pearl's account of actual causation has been formulated, let me turn consideration to Pearl's account of contributory causation. Clearly, this faces exactly the same problems as Pearl's account of actual causation – values of exogenous variables cannot be contributory causes. For, both accounts share the same definitions (P CM), (P CB), (P NB), and definition (P CC) repeats the criterion (C1) and (C2) of definition (P AC). In essence $X = x$ is classified as a contributory cause of $Y = y$ if there is causal beam that fulfils clauses (C1) and (C2), though not in any natural beam with this property.
However, the natural extension of this account of contributory causation can also be easily constructed by taking over definitions (EP CM), (EP CB) and (EP NB) from above and then reformulating definition (P CC) analogously to (EP AC). With this, the new extended definition of contributory causation is the following:

Definition (Contributory Cause, Extended Pearl) **(EP CC)**

For extended causal model M, event $X = x$ is said to be a contributory cause of $Y = y$ in a state u if and only if there exists an extended causal beam M_u, but no extended natural beam, such that:

$Y_x = y$ in M_u, (C1)

and

$Y_{x'} \neq y$ in M_u for some $x' \neq x$. (C2)

As in the case of actual causation, an application of my extended account of contributory causation to an example situation is equivalent to an application of Pearl's original account plus the introduction of dummy variables for every exogenous variable in the corresponding causal model.

CONTRIBUTORY CAUSATION: HITCHCOCK'S ACCOUNT AND MY EXTENSION OF PEARL'S ACCOUNT APPLIED TO TWO EXAMPLES

Compare now Hitchcock's account of contributory causation for cases of symmetric over-determination with my extended version of Pearl's account. Admittedly, after the discussion of my examples 1 and 2 we have already come to the conclusion that Hitchcock's project of simplifying Pearl's account of actual causation fails in two respects: by first having a different extension than Pearl's and second by contradicting our pre-theoretic causal intuitions. And certainly I have motivated an understanding of contributory causation as being a weakened version of actual causation that applies to cases of symmetric over-determination.

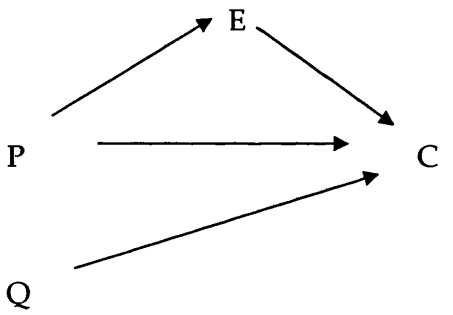
However, in strong contrast to Pearl, who offers a uniform account for actual and contributory causation, this is not the case with Hitchcock. For, his definition of a weakly active route in (H WAR) introduces the new notion of a redundancy set, by which the underlying rationale of the definition of an active route in (H AR) is given up: (H AR) allowed the freezing only of intermediate variables between cause variable X and effect variable Z, in order to screen off possible side effects of the intervention on X is. According to (H WAR), variables in a redundancy set $\{W_1,\ldots,W_n\}$ can be frozen at certain values independently from whether they are intermediate between X and Z or not. The redundancy condition only requires that the setting of the values $w_1,\ldots,w_n$ for the variables $W_1,\ldots,W_n$ does not alter the values of the variables in the previously selected route between X and Z.

For this reason, I consider a comparison of Hitchcock's account of contributory causation with my extended version of Pearl's account as very instructive. Moreover, this comparison also serves as a proving ground for my extended version of Pearl's account of actual and contributory causation. Lastly this comparison also illustrates how an application of my extended Pearl account proceeds with paradigmatic examples.

In the following I will consider natural generalizations of the above examples 1 and 2. These are achieved by introducing another causal mechanism that applies to the effect variable and works in parallel with the existing mechanisms. For the sake of simplicity I suppress a cover story and only focus on the formal structure of these examples.

The natural extension of my above example 1 is the following:

Example 3



This is given by the equations $f_E(p) = id_P$ and

$$f_C(p,e,q) = \begin{cases} 1, & \text{if } (E = P) \text{ or } (Q = 1) \\ 0, & \textit{otherwise} \end{cases}$$

with the actual state $P = 1$ and $Q = 1$,

yielding the solution $E = 1$ and $C = 1$.

Observe that now $P = 1$ at is not classified as an actual cause of $C = 1$ according to Hitchcock's account of actual causation with definitions (H AR) and (H AC) any more: Neither route (P,C) nor route (P,E,C) are

active, since in the actual model state with $Q = 1$ no intervention along the lines of definition (H AR) can change the value of C.

An application of Hitchcock's account of contributory causation with definitions (H WAR) and (H CC), though, yields the result that $P = 1$ is a contributory cause of $C = 1$. For, route (P,C) is weakly active according to Hitchcock's definition (H WAR):

Consider first route (P,E,C). Here, the only admissible choice of redundancy set $\{W_1,\ldots,W_n\}$ is $\{W_1,\ldots,W_n\} = \{Q\}$. But no matter whether we freeze variable Q at value $Q = 1$ or $Q = 0$, an intervention on P setting $P = 0$ does not bring about a change in the value of C. This is different for route (P,C). Here, the choice of redundancy set $\{W_1,\ldots,W_n\} = \{Q,E\}$ setting $Q = 0$ and $E = 1$ obviously fulfils the redundancy criterion. This combined freezing of Q at $Q = 0$ and E at $E = 1$ results in the solution $C = 0$, if we intervene at P setting $P = 0$. Thus a change in the value of P brings about a change in the value of C, rendering route (P,C) weakly active.

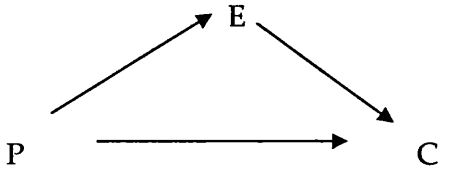
This result of an analysis by Hitchcock's definitions (H WAR) and (H CC) contradicts our causal intuition about this case, though. For, even allowing the contingency $Q = 0$, the value of variable C stays constant at $C = 1$, no matter whether $P = 1$ or $P = 0$. Hence we do not regard $P = 1$ as responsible for $C = 1$; neither in the actual scenario with $Q = 1$, nor in the contingent scenario with $Q = 0$.

This contradiction to our intuition, which the analysis by Hitchcock's account of contributory causation with (H WAR) and (H CC) yields, has a striking analogy. Applied to example 1, Hitchcock's account of actual causation with definitions (H AR) and (H AC) achieved an analogous verdict contradicting our causal intuition in that case.

Example 1 was the structurally simpler original of example 3 without the occurrence of variable Q. The functions for the other variables were

$f_E(p) = id_P$ and $f_C(p,e) = \begin{cases} 1, & \text{if } E = P \\ 0, & \text{otherwise} \end{cases}$, and the actual state $P = 1$ brought about the solution $E = 1$ and $C = 1$.

The graphical representation of example 1 was:



Hitchcock's account ruled of actual causation that route (P,C) was active, so that $P = 0$ was an actual cause of $C = 0$. But our causal intuition was already the same as it is now. The value of variable C stayed constant at $C = 1$, no matter whether $P = 1$ or $P = 0$, so that we did not regard $P = 1$ as responsible for $C = 1$. As a result, Hitchcock classified an event as an actual cause that was not a cause according to our pre-theoretic intuition, and with this I concluded that the account of actual causation with (H AR) and (H AC) was too wide in this respect.

Now, with regard to example 3, the analogous conclusion can be drawn that also Hitchcock's account of contributory causation with (H WAR) and (H CC) is too wide in certain respects, namely in characterizing events as contributory causes that are not according to our natural understanding. In this way, the weakening of the definition of an active route (H AR) by introducing the redundancy set $\{W_1,\ldots,W_n\}$ whose variables can be frozen at arbitrary values has not proven successful. Despite this admission of contingencies in the definition of a weakly active route (H WAR) Hitchcock's account of contributory causation with (H CC) can still not deal with cases like example 3 that are straightforward extensions of example 1.

Verify now that my extended version of Pearl's account of contributory causation yields the analogous result to the application of my extended version of Pearl's account of actual causation in example 1,[129] namely that $P = 1$ is neither an actual nor a contributory cause of $C = 1$:

For the present solution in the model with $P = 1$, $Q = 1$, $E = 1$ and $C = 1$, corresponding to state $F = 1$, $Q = 1$, we have, according to the definition of a causal beam (EP CB), the following choice for sustaining sets: [130]

$S_E = \{P\}$, $\underline{S}_E = \varnothing$ with either

$S_C = \{Q, P\}$, $\underline{S}_C = \{E\}$, or $S_C = \{E, P\}$, $\underline{S}_C = \{Q\}$, or

$S_C = \{Q, P, E\}$, $\underline{S}_C = \varnothing$.

Consider now the resulting natural beams for these choices in order to answer the question of actual causation, and also the associated non-natural causal beams with the same three choices of sustaining sets in order to investigate contributory causation.

First consider the natural beam with sustaining sets $S_E = \{P\}$ with $\underline{S}_E = \varnothing$, and $S_C = \{Q, P\}$ with $\underline{S}_C = \{E\}$. The resulting projection functions according to definition (EP CB) are

$$f_E^{F=1,Q=1}(p) = f_E(p), \text{ and } f_C^{F=1,Q=1}(q,p) = \begin{cases} 1, & \text{if } (P=1) \text{ or } (Q=1) \\ 0, & \text{otherwise} \end{cases}.$$

As required in definition (EP CB), both projection functions $f_E^{F=1,Q=1}$ and $f_C^{F=1,Q=1}$ are non-trivial. In this beam, an intervention on variable P in this natural beam setting $P = 0$ brings about the solution $E = 0$ and $C = 1$, since the value Q remains unaltered at $Q = 1$, showing that the value of C is not affected by this intervention.

---

[129] I noted earlier that an application of my extended version of Pearl's account of actual and contributory causation to an example situation is equivalent to an application of Pearl's respective original accounts plus the introduction of dummy variables for every exogenous variable in the corresponding causal model.

[130] Note that the choice of $S_C = \{Q\}$ with $\underline{S}_C = \{E, P\}$ is not admissible, because variable P may not occur in any contingency set $\underline{S}_C$, since it is member of the sustaining set of E.

There is only one non-natural causal beam with these sustaining sets, namely the beam with contingency subset $W_C = \{E\}$ setting E = 0.[131] This yields the projection

$f_C^{F=1,Q=1}(q,p) = \begin{cases} 1, & \text{if } (P=0) \text{ or } (Q=1) \\ 0, & \text{otherwise} \end{cases}$ and the same projection function as above for variable E, namely $f_E^{F=1,Q=1}(p) = f_E(p)$. Again, $f_C^{F=1,Q=1}$ is clearly non-trivial. Again, an intervention on P, now in this non-natural causal beam, setting $P = 0$ still brings about the solution $E = 0$ and $C = 1$, so that the value of C remains unaltered.

The second choice has the sustaining sets $S_E = \{P\}$ with $\underline{S}_E = \varnothing$, and $S_C = \{E, P\}$ with $\underline{S}_C = \{Q\}$. The resulting projection functions here would be $f_E^{F=1,Q=1}(p) = f_E(p)$ and the constant $f_C^{F=1,Q=1}(e,p) = 1$, the latter one obviously being trivial. That means that this choice of sustaining sets does not qualify as a natural beam.

Again, there is only one non-natural causal beam with these sustaining sets, this time with contingency subset $W_C = \{Q\}$ setting $Q = 0$.[132] This yields the projection

$f_C^{F=1,Q=1}(q,p) = \begin{cases} 1, & \text{if } (E=P) \\ 0, & \text{otherwise} \end{cases}$, being non-trivial in their arguments. However, again an intervention on P that sets $P = 0$, now in this new non-natural causal beam, does not alter the value of C.

The third choice of sustaining sets $S_E = \{P\}$ with $\underline{S}_E = \varnothing$, and $S_C = \{E, P, Q\}$ with $\underline{S}_C = \varnothing$ brings about the projections $f_E^{F=1,Q=1}(p) = f_E(p)$ and $f_C^{F=1,Q=1}(q,p) = f_C(p)$, so that the resulting natural beam is identical to the original model, keeping its non-trivial functions for E and C. Since $S_C = \{E, P, Q\}$ with $\underline{S}_C = \varnothing$, there only choice for a contingency subset is $W_C = \varnothing$, so that there are no non-natural beams with these sustaining sets.

---

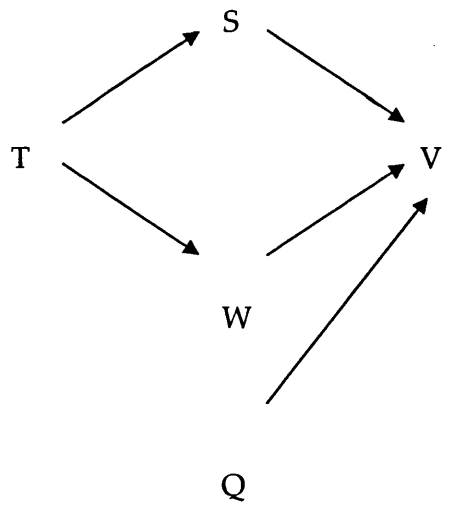
[131] For the other choices $W_C = \{E\}$ setting E = 1, or $W_C = \varnothing$, the solutions of the resulting beam coincide with the one of the above natural beam.

[132] For the other choices $W_C = \{Q\}$ setting Q = 1, or $W_C = \varnothing$, the solutions of the resulting beam again coincide with the one of the associated natural beam.

Taken together, this shows that according to my extension of Pearl's definition of actual causation by (EP AC) and of contributory causation by (EP CC), $P = 1$ is neither an actual nor a contributory cause for $C = 1$. For, there is neither a natural beam nor a non-natural causal beam in which an intervention on variable P changes the values of variable C. In this way, my extended version of Pearl's account is still in accord with our causal intuition and has passed this first test.

As the second and final point of comparison of Hitchcock's account of contributory causation with my extended version of Pearl's account, consider now the natural generalization of the above example 2 to a case of symmetric over-determination, again suppressing a cover story:

Example 4



This is given by the functions $f_S(t) = id_T$, $f_W(t) = id_T$ and

$$f_V(s,w,q) = \begin{cases} 0, & \text{if } S = 0 \text{ or } W = 0 \text{ or } Q = 0 \\ 1, & \text{otherwise} \end{cases}$$

and the actual state $T = 0$ and $Q = 0$,
yielding the solution $S = 0$, $W = 0$ and $V = 0$.

Observe that still $T = 0$ is not classified as an actual cause of $V = 0$ according to Hitchcock's account of actual causation with definitions (H AR) and (H AC): Neither route (T,S,V) nor route (T,W,V) are active, since in the actual model state with $Q = 0$ no intervention can change the value of V.

However, an application of Hitchcock's account of contributory causation for cases of symmetric over-determination, given by the definitions (H WAR) and (H CC), yields the result that $T = 0$ is a contributory cause of $V = 0$. For, both routes (T,S,V) and (T,W,V) are weakly active. In both cases choose the redundancy set $\{W_1,\ldots,W_n\} = \{Q\}$ setting $Q = 1$, such that an intervention on variable T setting $T = 1$ yields the solution $S = 1$, $W = 1$ and $V = 1$.

Verify now that my extended versions of Pearl's account of actual and contributory causation, given by definitions (EP NB) and (EP AC), or by (EP CB) and (EP CC) respectively, yields the same result that $T = 0$ is not an actual, but very well a contributory cause of $V = 0$:

For the present state of the model with $T = 0$, $Q = 0$, yielding the solution $S = 0$, $W = 0$, and $V = 0$, we have, according to the definition of a causal beam (EP CB), the following choice for sustaining sets: For variables S and W there is only one choice, which is $S_S = \{T\}$ with $S_S = \varnothing$, and $S_W = \{T\}$ with $S_W = \varnothing$. The resulting projection functions for these variables are hence identical to the functions in the original model with $f_S^{T=0,Q=0}(t) = id_T = f_S(t)$ and $f_W^{T=0,Q=0}(t) = id_T = f_W(t)$.

For variable V, there are seven possibilities for a choice of sustaining sets,

$S_V = \{S\}$ with $\underline{S}_V = \{W, Q\}$, or $S_V = \{W\}$ with $\underline{S}_V = \{S, Q\}$, or

$S_V = \{Q\}$ with $\underline{S}_V = \{S, W\}$, or $S_V = \{S, W\}$ with $\underline{S}_V = \{Q\}$, or

$S_V = \{S, Q\}$ with $\underline{S}_V = \{W\}$, or $S_V = \{W, Q\}$ with $\underline{S}_V = \{S\}$, or

$S_V = \{S, W, Q\}$ with $\underline{S}_V = \varnothing$.

Consider now the resulting natural beams for these choices, and verify that in none of them a change in the value of T can bring about a change in the value of V. For, observe that in the natural beams, in which Q is member of the sustaining set S_V, the actual value $Q = 0$ always ensures that $V = 0$ as well; and variable Q is obviously not affected by the intervention on T. If Q is a member of the contingency set $\underline{S}_V$ and thereby frozen at its actual value, this yields a constant projection function $f_V^{T=0,Q=0}(t) = 0$ which is obviously trivial and hence not admitted in the construction of a beam. As a result, there are no natural beams in which Q is contained in the contingency set S_V, and the only well-defined natural beams are innocuous, so that $T = 0$ does not actually cause $V = 0$, according to my definition (EP AC).

However, a non-natural causal beam exists with the required property, i.e. a beam in which an intervention on variable T setting $T = 1$ brings about $V = 1$, demonstrating that $T = 0$ is a contributory cause of $V = 0$, according to my definition (EP CC). Such a beam has to fulfill the following properties: First variable Q has to be member of contingency subset $W_V \subseteq \underline{S}_V$, and second variable S or W have to be contained in the sustaining set S_V, so that an intervention on variable T can propagate to variable V. A beam meeting these requirements is given by making the following choices of sustaining sets: Choose $S_V = \{S, W\}$ with $\underline{S}_V = \{Q\}$, and for variables S and W follow the only possibility $S_S = \{T\}$ with $\underline{S}_S = \varnothing$, and $S_W = \{T\}$ with $\underline{S}_W = \varnothing$. In this way, the projection functions with $f_S^{T=0,Q=0}(t) = id_T = f_S(t)$ and $f_W^{T=0,Q=0}(t) = id_T = f_W(t)$ result for variables S and W. The projection function of variable V is given by $f_V^{T=0,Q=0}(s,w,q) = \begin{cases} 0, & \text{if } S = 0 \text{ or } W = 0 \\ 1, & \text{otherwise} \end{cases}$.

Obviously, now an intervention on variable T setting $T = 1$ yields the solution $S = 1$, $W = 1$ and the desired result $V = 1$.[133]

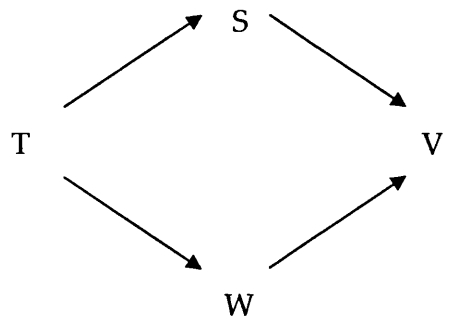
---

[133] The same would hold for the non-natural causal beams resulting from the following combinations of sustaining sets S and contingency subsets W:
$S_V = \{S\}$ with $\underline{S}_V = \{W, Q\} = W_V$, setting $W = 1$ and $Q = 1$;

With this it has been demonstrated that according to my extended versions of Pearl's account of actual and contributory causation, given by definitions (EP NB) and (EP AC), or by (EP CB) and (EP CC) respectively, $T = 0$ is not an actual, but very well a contributory cause of $V = 0$.

This verdict coincides with our natural understanding of the situation. For under the assumption of $Q = 1$, clearly the net effect of $T = 0$ is $V = 0$. More elaborately, we do not regard $T = 0$ as responsible for $V = 0$ in the actual scenario given by $Q = 0$; that is why we do not understand $T = 0$ as actually causing $V = 0$. However, in the contingent scenario given by $Q = 1$, we regard $T = 0$ as being very well responsible for $V = 0$; and that is the reason why we understand $T = 0$ as causally contributing to $V = 0$.
As a result, my extended version of Pearl's account of actual and contributory causation, given by definitions (EP NB) and (EP AC), or by (EP CB) and (EP CC) respectively, has passed its second test.

Let me lastly come back to the analogy with my example 2, which was the structurally simpler original of example 4 without the occurrence of variable Q. The functions for the other variables were $f_S(t) = id_T$, $f_W(t) = id_T$, and $f_V(s,w) = \begin{cases} 1, & \text{if } S=1 \text{ and } W=1 \\ 0, & \text{otherwise} \end{cases}$. The actual state of the model was $T = 0$ with the resulting solution $S = 0$, $W = 0$ and $V = 0$. The associated graphical representation of example 2 was:



and analogously $S_V = \{W\}$ with $\underline{S}_V = \{S, Q\} = W_V$, setting $S = 1$ and $Q = 1$.

An application of Hitchcock's account of actual causation yielded the result that neither route (T, S, V) nor route (T, W, V) were active, so that $T = 0$ was not an actual cause of $V = 0$ according to Hitchcock.

Observe that this gap that existed in example 2 between the verdict of Hitchcock's account for actual causation and our intuition about actual causation is closed for contributory causation in example 4. Hitchcock's account of contributory causation, given by definitions (H WAR) and (H CC), can detect $T = 0$ as being a contributory cause of $V = 0$.

REASONS FOR THE FAILURE OF HITCHCOCK'S ACCOUNT OF ACTUAL AND CONTRIBUTORY CAUSATION

In the following, I investigate the differences between Hitchcock's account of contributory causation, given by definitions (H WAR) and (H CC), and his account of actual causation employing definitions (H AR) and (H AC). In the above example 4 we noticed that Hitchcock's account of contributory causation is in a certain way an improvement over his account of actual causation, since there an analysis with (H WAR) and (H CC) delivered a result in accordance with our intuition. However, example 3 showed that there are still cases where (H WAR) and (H CC) still yield analogous results to (H AR) and (H AC) that contradict our intuition. Understandably, I am particularly interested in why Hitchcock's account of contributory causation still fails, despite being an improvement over his account of actual causation.

Clearly, the core of Hitchcock's accounts of actual and of contributory causation lies in his definitions of an active route (H AR) or of a weakly active route, respectively. The decisive deviation in (H WAR) from (H AR) is the possibility of choosing the set of variables $\{W_1,\ldots,W_n\}$, whose values are arbitrarily fixed, independently from whether they are intermediate between cause variable X and effect variable Z or not. The redundancy condition only requires that the setting of the values $w_1,\ldots,w_n$ for the variables $W_1,\ldots,W_n$ does not alter the values of the variables in the previously selected route $(X,Y_1,\ldots,Y_n,Z)$ for the actual solution in the model.

Another possible weakening of Hitchcock's definition of an active route (H AR) in order to cover cases of symmetric over-determination would, though, be the following: Take $\{W_1,\ldots,W_n\}$ to be a set of intermediate variables between X and Z such that an arbitrary fixation of these variables W_i at the values w_i does not alter the value of $Y_1,\ldots,Y_n$ and Z. Observe though that this formulation could not account for symmetric over-determination, for it would still fail for obvious reasons for my example 3: Route (P,C) would still be weakly active in this understanding, because we could still choose the redundancy set as $\{W_1,\ldots,W_n\} = \{Q,E\}$ and set $Q = 0$ and $E = 1$.

Furthermore, we could also add a freezing of intermediate variables at their actual values to Hitchcock's definition (H WAR). $\{W_1,\ldots,W_n\}$ would be taken to be an arbitrary set of variables disjoint from the set of intermediate variables between X and Z such that an arbitrary fixation of these variables W_i at the values w_i does not alter the value of $Y_1,\ldots,Y_n$ and Z. And the variables that are intermediate between X and Z but do not belong to the route $(X,Y_1,\ldots,Y_n,Z)$ would be frozen at their actual values. This formulation would fail in example 4, though, for the same reasons as (H AR) failed in example 2: Neither route (T, S, V) nor route (T, W, V) would be weakly active in this understanding, since for route (T, S, V) the intermediate variable W would be frozen at is actual value, and for route (T, W, V) the same would hold for variable S.

Both these variants would conserve the rationale underlying Hitchcock's account of actual causation: isolate causal processes by freezing intermediate variables between X and Z not belonging to the route between X and Z in question at their actual values, in this way screening off unwanted side effects of the intervention on the cause variable X.

Hitchcock's favored definition of a weakly active route (H WAR) deviates from this rationale. Instead (H WAR) is a huge step in the direction of Pearl's account by reflecting a notion of sustaining the actual values along

the route between X and Z against contingencies that are introduced by the redundancy set.

Obviously, definition (H WAR) of Hitchcock's fares better in comparison to my extended version of Pearl's account on contributory causation than his definition (H AR) compares to my extended version of Pearl's account on actual causation. However, (H WAR) is not a natural generalization of (H AR), but instead an ad hoc concession to deal with cases of a special structure, i.e. symmetric over-determination, while also incorporating a certain aspect of sustenance. In my view, this strongly suggests that Hitchcock's original idea underlying (H AR) and (H AC) to identify causal processes in a situation with concrete or graphic routes in the causal model of this situation heads in the wrong direction.

This conjecture gets more evidence when (H AR) is modified along the lines of (H WAR). The straightforward reduction of (H WAR) to deal with cases of actual causation would namely be the following:

Definition (Strongly Active Route, Modified Hitchcock) **(MH SAR)**

The route $(X,Y_1,\ldots,Y_n,Z)$ is *strongly active* in a solution $\boldsymbol{R}(\boldsymbol{V})$ of the causal model $\boldsymbol{M} := \langle \boldsymbol{V}, \boldsymbol{E} \rangle$ if and only if there exists a set (possibly empty) of variables $\{W_1,\ldots,W_m\}$ in $\boldsymbol{V} \setminus \{X,Y_1,\ldots,Y_n,Z\}$, such that the following holds:

The value z of Z depends counterfactually upon the value x of X within the resulting solution $\boldsymbol{R}'(\boldsymbol{V})$ of the new system of equations $\boldsymbol{E}'$, constructed from $\boldsymbol{E}$ as follows:

For each W_i, replace the equation E_{W_i} for W_i with the new equation $E'_{W_i} := W_i = a_i$ that sets W_i constant to its value a_i in the solution $\boldsymbol{R}(\boldsymbol{V})$.

Under this modified account, the above example 2 would be analyzed differently yielding the same result as an application of my extended Pearl account of actual causation, namely identifying $T = 0$ as an actual

cause of $V = 0$: In the actual state $T = 0$ of the model, resulting in $S = 0$, $W = 0$ and $V = 0$, both routes (T, S, V) and (T, W, V) are strongly active according to definition (MH SAR), since one can choose the redundancy set to be $\{W_1,\ldots,W_m\} = \varnothing$ in both cases. This way, the original functions of the model are preserved, and an intervention on T setting $T = 1$ brings about $S = 1$, $W = 1$, and in turn $V = 1$, so that according to (MH SAR) $T = 0$ is an actual cause of $V = 0$.

However, this modified account would still have the same shortcoming as Hitchcock's original account in analyzing example 1 as containing a case of genuine actual causation: For, observe that in the actual state $P = 1$, yielding the solution $E = 1$ and $C = 1$, route (P, C) is still strongly active according to definition (MH SAR), ruling $P = 1$ to be an actual cause of $C = 1$: Choose redundancy set $\{W_1,\ldots,W_m\} = \{E\}$, and freeze E at its actual value $E = 1$, so that an intervention on P setting $P = 0$ brings about $C = 0$, because variables E and P have different values.

These considerations can only lead to the result that the rationale underlying Hitchcock's accounts of actual and contributory causation, given by (H AR) and (H WAR), namely to identify a causal process in a situation with a route in the causal model, seems to be misguided. The notion of a causal process linking a cause with its effect in a situation requires a structurally more complicated construction to be formalized. I maintain that such a construction is given by Pearl's definition of a causal beam that also captures a notion of sustenance. My slight extension of Pearl's definition yielding my extended causal beams conserves these two underlying notions. And so far, the verdicts my extended account delivers are in accord with our causal intuition.

INCLUSIVE TOKEN CAUSATION: WOODWARD'S ACCOUNT APPLIED TO MY FOUR EXAMPLES

Let us now briefly investigate how Woodward's account of token causation fares with regard to the above examples. We observed in Chapter 1 that Woodward does not distinguish between actual and contributory causation. For Woodward, there is only one notion of token causation, and he attempts to capture this by his definition (W TC).

Woodward understands token causation to be an inclusive notion of singular event causation that also applies to cases of symmetric over-determination. However, his definition of token causation (W TC) is equivalent to Hitchcock's definition of contributory causation (H CC) employing the notion of a weakly active route defined by (H WAR). The difference from Hitchcock is that Woodward does not limit the extension of his definition (W TC) to cases of symmetric over-determination, as Hitchcock does, but that Woodward takes his definition to cover all cases of singular event causation.

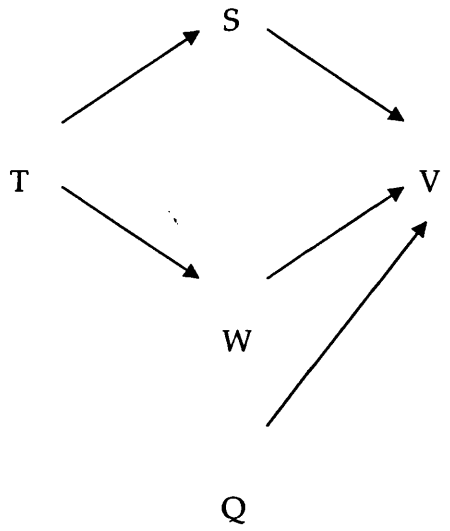
Since Woodward's definition (W TC) is equivalent to Hitchcock's definition of contributory causation (H CC) employing the definition of a weakly active route (H WAR), we can immediately verify the verdicts of Woodward's account when it is applied to my examples 3 and 4 above.

With regard to my first two examples, one has to keep in mind that Hitchcock's definition of an active route (H AR) was a special case of a weakly active route according to (H WAR), so that for Hitchcock actual causation was a special case of contributory causation.

In the following, I give a detailed description of how an application of Woodward's account of token causation given by (W TC) to my four examples proceeds, dealing with the examples in reversed order.

Application of Woodward's Definition (W TC) to Example 4

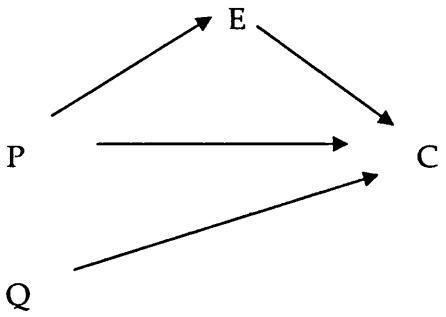
Woodward's definition (W TC) successfully passes the test posed by example 4. This example was given by the functions $f_S(t) = id_T$, $f_W(t) = id_T$, and $f_V(s,w,q) = \begin{cases} 0, & \text{if } S=0 \text{ or } W=0 \text{ or } Q=0 \\ 1, & \text{otherwise} \end{cases}$. The actual state of the model with $T = 0$ and $Q = 0$ resulted in the solution $S = 0$, $W = 0$ and $V = 0$. The graphical representation was:



Under an analysis with Woodward's definition of token causation (W TC), $T = 0$ is a token cause of $V = 0$: For, my discussion of how Hitchcock's account of contributory causation with definitions (H CC) and (H AR) is applied to example 4 has shown that both routes (T,S,V) and (T,W,V) were weakly active. The redundancy set was for both routes chosen as $\{W_1,\ldots,W_n\} = \{Q\}$, setting $Q = 1$, such that an intervention on variable T setting $T = 1$ yields the solution $S = 1$, $W = 1$ and $V = 1$. As a result, $T = 0$ was a contributory cause of $V = 0$ according to Hitchcock's (H CC). With this, $T = 0$ clearly is a token cause of $V = 0$ according to Woodward's definition (W TC), since as noted above (W TC) is equivalent to (H CC). And this verdict accords to our intuition for this case that $V = 0$ is a net effect of $T = 0$.

Application of Woodward's Definition (W TC) to Example 3

However, example 3 is a clear counter-example to Woodward's account. This example had the graphical representation:



It was given by functions $f_E(p) = id_P$ and $f_C(p,e,q) = \begin{cases} 1, & \text{if } (E = P) \text{ or } (Q = 1) \\ 0, & \text{otherwise} \end{cases}$.

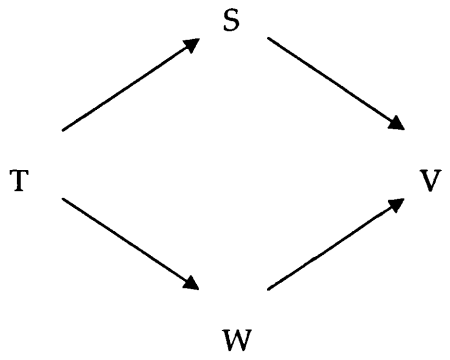
The actual state of the model was $P = 1$ and $Q = 1$, which resulted in the solution $E = 1$ and $C = 1$.

An application of Woodward's definition of token causation (W TC) now yields the result that $P = 1$ is a token cause of $C = 1$: For, route (P,C) was weakly active according to Hitchcock's definition (H WAR). With the redundancy set $\{W_1,\ldots,W_n\} = \{Q,E\}$ setting $Q = 0$ and $E = 1$, an intervention on P setting $P = 0$ resulted in the solution $C = 0$, i.e. a change in the value of P brought about a change in the value of C. As a result, $P = 1$ was a contributory cause of $C = 1$ according to Hitchcock's (H CC). And with this $P = 1$ clearly is a token cause of $C = 1$ according to Woodward's definition (W TC), since as noted above (W TC) is equivalent to (H CC).

But, as outlined above, this result contradicts our pre-theoretic intuition, since even assuming that $Q = 0$, the value of variable C stays constant at $C = 1$ no matter whether $P = 1$ or $P = 0$. An analogous conclusion as in Hitchcock's case can be drawn, namely that Woodward's account of token causation given by (W TC) is also too wide in certain respects. It characterizes events as token causes that are not according to our natural understanding.

Application of Woodward's Definition (W TC) to Example 2

Example 2 was the basis out of which my example 4 was generated by adding variable Q to reach a case of symmetrical over-determination for variable V. Example 2 was given by the functions $f_S(t) = id_T$, $f_W(t) = id_T$, and $f_V(s,w) = \begin{cases} 1, & \text{if } S=1 \text{ and } W=1 \\ 0, & \text{otherwise} \end{cases}$. The actual state of the model was $T = 0$ with the resulting solution $S = 0$, $W = 0$ and $V = 0$. The associated graphical representation was:



An application of Hitchcock's definitions (H AC) of actual causation and (H AR) of an active route yielded the verdict that neither route (T, S, V) nor route (T, W, V) were active. As a result, $T = 0$ was not an actual cause of $V = 0$ according to Hitchcock.

However, under an analysis with Woodward's definition of token causation (W TC), $T = 0$ is a token cause of $V = 0$: For, observe that analogously to my discussion of how Woodward's account of token causation is applied to example 4, both routes (T,S,V) and (T,W,V) are again weakly active. Choose the redundancy set for both routes as $\{W_1,\ldots,W_n\} = \varnothing$,[134] such that an intervention on variable T setting $T = 1$ in both cases yields the solution $S = 1$, $W = 1$ and $V = 1$. As a result, both routes (T,S,V) and (T,W,V) are weakly active according to Hitchcock's

[134] Remember that Hitchcock's definition of a weakly active route clearly allowed empty redundancy sets. In this case, this setting of $\{W_1,\ldots,W_n\} = \varnothing$ results in freezing no variables while intervening on T.

definition (H WAR), so that $T = 0$ is a contributory cause of $V = 0$ according to Hitchcock's (H CC). With this $T = 0$ clearly is a token cause of $V = 0$ according to Woodward's definition (W TC), since as noted above (W TC) is equivalent to (H CC).

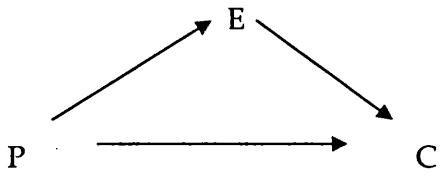
Our intuition for this case ruled that $V = 0$ is a net effect of $T = 0$. Hence, analogously to example 4, Woodward's account of token causation given by (W TC) is once more in accord with our pre-theoretic intuition. However, this time the verdict of Woodward's account of token causation deviates from the one that Hitchcock's account of actual causation delivered for this example.

This divergence in verdict gets more plausible by remembering that Hitchcock distinguished between actual and contributory causation. Examples 3 and 4 are cases of symmetric over-determination, for which Hitchcock specifically developed his account of contributory causation with (H CC). For these two examples Woodward's account reached the same verdicts, since his definition of token causation (W TC) is equivalent to Hitchcock's definition of a contributory cause. However, whereas Woodward's notion of token causation is an inclusive one covering all cases of single event causation, Hitchcock's notion of contributory causation was not: For Hitchcock, actual causes certainly were a special case of contributory causes; however, the conditions an actual cause had to fulfill were much stricter than the ones expected from a contributory cause.

Viewed from this angle, example 2 gives a clear case in which Woodward's account of token causation is superior to Hitchcock's account of actual causation. I leave it to the reader, however to judge whether this counts as evidence in favor of an inclusive understanding of token causation.

Application of Woodward's Definition (W TC) to Example 1

Example 1 was the basis out of which my example 3 was generated by adding variable Q to reach a case of symmetrical over-determination for variable C. Example 1 had the graphical representation



and was given by the functions $f_E(p) = id_P$ and

$f_C(e,p) = \begin{cases} 0, & \text{if E and P have different values} \\ 1, & \text{if E and P have the same value} \end{cases}$. The actual state of the model was $P = 1$, so that the resulting solution was $E = 1$ and $C = 1$.

An application of Hitchcock's account of actual causation given by definitions (H AC) and (H AR) yielded the verdict that $P = 0$ is an actual cause of $C = 0$, since the route (P,C) is active.

As mentioned above, every active route r according to definition (H AR) is in particular weakly active according to definition (H WAR): In the application of definition (H WAR), choose the redundancy set as the set of intermediate variables not belonging to the route r, and freeze these variables at their actual values. As a result, every actual cause according to (H AC) is in particular a contributory cause along the lines of (H CC) for Hitchcock.

It is instructive to verify in the following that route (P,C) is weakly active according to (H WAR), too. Choose the redundancy set $\{W_1,\ldots,W_n\} = \{E\}$ and freeze variable E at its actual value $E = 1$. Then an intervention on P setting $P = 0$ results in the solution $C = 0$, i.e. a change in the value of P brings about a change in the value of C. As a result, $P = 1$ is a contributory cause of $C = 1$ according to Hitchcock's (H CC).

Since as noted above (W TC) is equivalent to (H CC), with this, $P = 1$ clearly is a token cause of $C = 1$ according to Woodward's definition

(W TC), too. And as outlined above, too, this result contradicts our pre-theoretic intuition, since the value of variable C stays constant at $C = 1$, no matter whether $P = 1$ or $P = 0$.

Hence, this result reinforces the conclusion that could already be drawn in example 3: Woodward's account of token causation given by (W TC) is too wide in certain respects, namely by characterizing events as token causes that are not according to our natural understanding.

REASONS FOR THE FAILURE OF WOODWARD'S ACCOUNT AND AN EXTENSION OF PEARL'S ACCOUNT TO COVER INCLUSIVE TOKEN CAUSATION

Let me briefly summarize what has been achieved by the application of Woodward's account of token causation to my four examples in the last section of this chapter:

I could verify that this inclusive understanding of token causation of Woodward's marks an improvement over Hitchcock's separated notions of actual and contributory causation: In example 2, Woodward's account with (W TC) delivered a result that was in accord with our causal intuition; Hitchcock's account of actual causation with (H AC) and (H AR) previously failed with this example. However, Woodward's account with (W TC) still analyzed the examples 1 and 3 in a way yielding results that contradicted our causal intuition. And this should give sufficient evidence to reject his account.

In the last but one section of this chapter, I maintained that the reason for the failure of Hitchcock's account of actual and contributory causation was the fixation on causal routes. Hitchcock understood causal routes in a model to be the formalization of a causal process that links a cause with its effect.

The same rationale underlies Woodward's account of token causation, since he adopts Hitchcock's definition (H CC). And this is the reason for the failure of Woodward's account in examples 1 and 3: Identifying causal processes with routes r in a model between cause and effect variables A and Z, respectively, leads to a freezing of intermediate variables Y that do not belong to r at their actual values y. If these variables Y are involved in

sustaining the actual value of the effect variable Z, then a freezing of Y at their actual values distorts the interaction of the mechanisms that are at work in the modeled situation. This is exactly what happens in examples 1 and 3. The route in question is (P,C) with the intermediate variable *E*. Variable *E* is involved in sustaining the actual value of the effect variable *C*, hence a freezing of E at its actual value distorts the interaction of the mechanisms in the modeled situation.

This failure of Woodward's account of token causation and the fact that it fails because of the same reason as Hitchcock's account of actual and contributory causation gives a final justification to the way in which I discuss Woodward's account in this thesis. I decided to give only a short presentation and discussion of Woodward's account since it adopted one of Hitchcock's definitions, where Hitchcock's account was the earlier one and in strong need of conceptual clarification. However, now it has also turned out that Woodward's account does not offer us huge advantages over Hitchcock's account in terms of its analytical abilities. Where Hitchcock's account failed in three examples out of four, Woodward's account only failed in two.

The main improvement of Hitchcock's account that Woodward offered was the uniform treatment of all cases of singular event causation by the same definition (W TC) with an underlying understanding of inclusive token causation. However, in the following I argue that such a uniform treatment of all cases of singular event causation can also be achieved by my extension of Pearl's account, if one makes a minimal modification in my definition of extended contributory causation (EP CC).

Let me start with some broad considerations about unified token causation. Certainly such an understanding of token causation as a unified notion comprising both actual and contributory causation has appeal. After all, in both cases we had the same basic rationale of a causal process linking token cause and token effect underlying Hitchcock's as well as

Pearl's account. The difference between actual and contributory causation for Hitchcock and Pearl was located in the kind of contingencies in which a causal process operates. For actual causation this process had to operate in the very circumstances prevailing in the situation to be modeled: both Hitchcock and Pearl froze the variables not immediately involved in the process at their actual values. For contributory causation, they allowed the process to operate in non-actual circumstances that correspond to freezing variables not immediately involved in the process at arbitrary values, with the only constraint being that this did not disturb the actual value of the effect variable.

Sharing this underlying rationale, the main difference in Hitchcock's and Pearl's accounts resulted from a different formalization of a process: Hitchcock identified a process with a route in the causal model whereas Pearl equated a process to a projection of the causal model. In Pearl's projections, his so-called beams, the resulting projection functions were still allowed to have more than one argument. For Hitchcock, the functions associated to the variables in a route were effectively only allowed one argument, because all other variables were frozen.

Putting these considerations aside, though, we noticed in chapter 1 that such a unified understanding of token causation was not a major breakthrough. After all, Woodward's definition (W TC) was equivalent to Hitchcock's definition of contributory causation employing Hitchcock's definition of a weakly active route (H WAR). And Hitchcock's definition of an active route (H AR) was a special case of a weakly active route, so that for Hitchcock actual causation was a special case of contributory causation. With his preliminary definition of token causation (W TCP) Woodward even gave a definition that was equivalent to Hitchcock's definition of actual causation (H AC).

Admittedly, this was though only a heuristic step for Woodward in order to motivate his final definition of token causation (W TC). Moreover, to

make another admission, neither Woodward nor Hitchcock used the term 'contributory causation'. Hitchcock developed his weakly active routes in order to deal with cases of symmetric over-determination, but did not really distinguish two notions of token causation. I borrowed the term 'contributory causation' from Pearl for my description of Hitchcock's account for the mere reason that I found this to be a practical way of talking.

Even Pearl does not really distinguish between actual and contributory causation as being two kinds of token causation. For, natural beams defined by (P NB) are just a special case of causal beams defined by (P CB), namely the ones in which all the variables in the complements $\underline{S}$ of the sustaining sets S are frozen at their actual values. Pearl's way of formulating actual and contributory causation as being mutually exclusive concepts is hence completely arbitrary. For, delete the phrase 'but no natural beam' from Pearl's definition of contributory causation:

Definition (Contributory Cause, Pearl) **(P CC)**

An event $X = x$ is said to be a contributory cause of $Y = y$ in a state $\boldsymbol{u}$ if and only if there exists a causal beam $\boldsymbol{M_u}$, *but no natural beam*, such that:

$$Y_x = y \text{ in } \boldsymbol{M_u}, \qquad \text{(C1)}$$

$$\text{and } Y_{x'} \neq y \text{ in } \boldsymbol{M_u} \text{ for some } x' \neq x. \qquad \text{(C2)}$$

The result is an inclusive notion of contributory causation that comprises actual causation as a special case, since as said every natural beam is in particular a causal beam. With this modification, one could rightfully call this a definition of token causation containing actual causation as a special case. Hence, I would finally like to suggest this modification for my definition of extended contributory causation in order to yield an inclusive definition of extended token causation:

Definition (Token Cause, Extended Pearl) **(EP TC)**

An event $X = x$ is said to be a token cause of $Y = y$ in a state $\boldsymbol{u}$ if and only if there exists a causal beam $\boldsymbol{M_u}$ such that:

$$Y_x = y \text{ in } \boldsymbol{M_u}, \qquad \text{(C1)}$$

$$\text{and } Y_{x'} \neq y \text{ in } \boldsymbol{M_u} \text{ for some } x' \neq x. \qquad \text{(C2)}$$

Apparently, with this one reaches a uniform treatment of all cases of singular event causation by one underlying notion of inclusive token causation. Definition (EP TC) expresses the canonical unification of an extended actual and an extended contributory cause according to definitions (EP AC) and (EP CC), respectively.

However, in my understanding this issue of a inclusive understanding of token causation applying to all cases of singular event causation simultaneously was more a terminological problem than an analytical one. The important fact to remember is that my extension of Pearl's account has analyzed all my paradigmatic examples successfully.

MY EXTENSION OF PEARL'S ACCOUNT APPLIED TO CLASSIC CASES OF INTRANSITIVITY, PREEMPTION, AND SYMMETRIC OVER-DETERMINATION

As a final proving ground for my extension of Pearl's account of actual and contributory causation I briefly demonstrate that it does not fare worse than Hitchcock's original account in dealing with the three major problems of Lewis's classic counterfactual account: the intransitivity of causation, cases of preemption, and cases of symmetric over-determination.

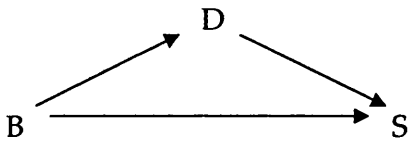
Application of (EP AC) to a Classic Example Featuring Intransitivity

Let me start with the intransitivity of actual causation. Admittedly, I do not have a formal proof that my account does not fare worse than Hitchcock's regarding this problem. However, observe that in order to motivate the superiority of his own account of actual causation over Lewis's classic counterfactual account Hitchcock himself settled for two things: The first was a remark that his account of actual causation did not stipulate transitivity. The second was a discussion of two examples, which illustrated a failure of composition of actual causation, together with the demonstration that their analysis by his account delivered a verdict in accord with our causal intuition.

In the same spirit, let me first remark that my extension of Pearl's account apparently nowhere stipulates or assumes transitivity of actual causation. Second, let me demonstrate that an application of my extension of Pearl's account of actual causation to Hitchcock featured two examples delivers

analogous verdicts as Hitchcock's own account. These two examples of Hitchcock's shared a common structure; in essence they just contained three events *c*, *d*, and *e* such that *c* was an actual cause of *d* in our understanding, *d* an actual cause of *e*, but *c* failed to be an actual cause of *e* in our intuition. I settle for a discussion of Hitchcock's favored example 'Boulder' here because the analysis of his other example 'Dog Bite' proceeds in exactly the same way.[135]

The graphical representation of the 'Boulder' example was:



The functions were $f_D = id_B$ and $f_S = \begin{cases} 0, & \text{if B = 1 and D = 0} \\ 1, & \text{otherwise} \end{cases}$, so that the model state $B = 1$ yielded the actual solution $D = 1$ and $S = 1$.

In Hitchcock's analysis, route (B,D) was active, and the same applied to route (D,S), given the actual state of the model with $B = 1$. However, neither of the routes (B,S) and (B,D,S) were active, since an intervention on B setting $B = 0$ does not alter the value of S, no matter whether D is frozen at its actual value $D = 1$ or not. According to (H AC), with this $B = 1$ was an actual cause of $D = 1$, $D = 1$ was an actual cause of $S = 1$, but $B = 1$ was not an actual cause of $D = 1$, thus showing a failure of composition.

According to my extended definition of a causal beam (EP CB), there is only one admissible choice of sustaining sets, namely $\mathcal{S}_D = \{B\}$ with $\underline{S}_D = \emptyset$ and $\mathcal{S}_S = \{B, D\}$ with $\underline{S}_S = \emptyset$. Hence, there is only one natural beam to consider, and this beam is identical to the original model with the original functions f_D and f_S as projection functions $f^{B=1}{}_D$ and $f^{B=1}{}_S$. Obviously, in this beam an intervention on variable B setting $B = 0$ changes the value of variable D to $D = 0$, so that according to definition (EP AC) $B = 1$ is an extended actual cause of $D = 1$. Analogously, an intervention on variable D

[135] For Hitchcock's original example 'Dog Bite' cf. [cH 01], pp. 290-295.

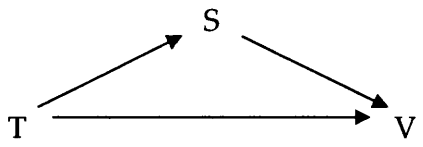
setting $D = 0$ changes the value of variable S to $S = 0$, so that also $D = 1$ is an extended actual cause of $S = 1$. However, an intervention on variable B setting $B = 0$ does not affect the value of variable S, so that $B = 1$ is not an extended actual cause of $D = 1$.

With this, an application of my account reaches exactly the same verdict as an application of Hitchcock's account to this example, showing that my extended actual causation as defined by (EP AC) is not a transitive relation either in general.

Application of (EP AC) to a Classic Example Featuring Preemption

Second, let me discuss preemption. Again, I do not give a formal proof that my account of actual causation does not fare worse than Hitchcock's own regarding this problem. Instead I settle again for demonstrating that an application of my account to Hitchcock's featured examples delivers the same verdicts as his own account.

This time, Hitchcock only discusses one example in detail that contains a cause preempting another one, and that is his example 'Backup'.[136] The causal model for the example 'Backup' was given by the functions $f_S = \underline{sig}_T$, with $\underline{sig}_Y = 1$, if $Y = 0$ and $\underline{sig}_Y = 1$, if $Y \neq 0$, for an arbitrary variable Y, and by $f_V = \begin{cases} 0, & \text{if } T = 0 \text{ and } S = 0 \\ 1, & \text{otherwise} \end{cases}$. The model state $T = 1$ yielded the actual solution $S = 0$ and $V = 1$, and the graphical representation was:



An application of Hitchcock's account showed that route (T, V) is active, since by freezing S at its actual value $S = 0$, an intervention on T setting

[136] Cf. [cH 01], pp. 287-288.

$T = 0$ changes the value of V. In this way, Hitchcock could analyze this example correctly according to our intuition ruling that $T = 1$ is an actual cause of $V = 1$, despite preempting $S = 1$ that would otherwise bring about $V = 1$ itself.

An application of my extended version of Pearls account of actual causation proceeds in the following way: According to (EP CB), there are two admissible choices of sustaining sets, namely $S_S = \{T\}$ with $\underline{S}_S = \emptyset$ for variable S, and $S_V = \{T, S\}$ with $\underline{S}_V = \emptyset$, or $S_V = \{T\}$ with $\underline{S}_V = \{S\}$ for variable V, so that there are two natural beams to consider.[137]

Consider the beam with sustaining sets $S_S = \{T\}$, $\underline{S}_S = \emptyset$, and $S_V = \{T\}$, $\underline{S}_V = \{S\}$. Its projection functions are $f^{\beta=1}_S = \underline{sig}_T = f_S$, and

$f^{\beta=1}_V = \begin{cases} 0, & \text{if } T = 0 \\ 1, & \text{if } T = 1 \end{cases}$, since variable S is frozen at its actual value $S = 0$.

Clearly, in this beam an intervention on variable T setting $T = 0$ changes the value of variable V to $V = 0$. According to my definition (EP AC), $T = 1$ is thus an extended actual cause of $V = 1$.

As a result, the analysis of this example by my extended version of Pearl's account gives the same result as an application of Hitchcock's account.

Application of (EP CC) to a Classic Example Featuring Symmetric Over-Determination

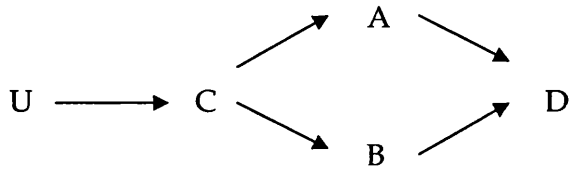
Let me lastly discuss the problem of symmetric over-determination. Here, Hitchcock himself does not really discuss an example in detail that contains an effect being caused in a parallel way by more than one cause.[138] Instead, let me conclude this chapter with the discussion of Lewis's classic example 'Firing Squad'.

---

[137] The choice of $S_V = \{S\}$ with $\underline{S}_V = \{T\}$ for the sustaining set of variable V is not admissible, since variable T is contained in the sustaining set S_S of variable S.

[138] As already mentioned, Hitchcock's main concern in [cH 01] is to show the intransitivity of actual causation.

The causal model for this example was given by the graph



with the structural equations $C = U$, $B = C$, $A = C$ and $D = \max\{A,B\}$, and the actual state $U = 1$, yielding the solution $C = 1$, $A = 1$, $B = 1$, and in turn $D = 1$. As mentioned in the introduction of this thesis, Lewis's account could not answer the question whether the shooting of rifleman B, formalized by $B = 1$ is a contributory cause of the death of the prisoner, formalized by $D = 1$, which is why he tried to find a way out by proclaiming that he had no intuition in this case.

I have shown, though, that Hitchcock's account could analyze this example successfully, with the upshot that $B = 1$ is not an actual cause of $D = 1$, but that route (B, D) is weakly active. In this way, the verdict of Hitchcock's account coincided with my intuition, which I consider to be the natural one, namely that the shooting by the rifleman is very well a contributory cause of the death of the prisoner.

For an application of my extended Pearl account of actual and contributory causation, respectively, observe that according to definition (EP CB) we have the following choice of sustaining sets: $S_C = \{U\}$ with $\underline{S}_C = \emptyset$ for variable C, $S_A = \{C\}$ with $\underline{S}_A = \emptyset$ for variable A, $S_B = \{C\}$ with $\underline{S}_B = \emptyset$ for variable B, and for variable D $S_D = \{A, B\}$ with $\underline{S}_D = \emptyset$, or $S_D = \{A\}$ with $\underline{S}_D = \{B\}$, or $S_D = \{B\}$ with $\underline{S}_D = \{A\}$.

Verify first that $B = 1$ is not an extended actual cause for $D = 1$: Observe that the latter two choices of sustaining sets for variable D are not admissible for the construction of a natural beam, because the resulting projection function $f^{U=1}{}_D$ would be constant at value $f^{U=1}{}_D = 1$ and hence trivial. In the only remaining natural beam with sustaining set $S_D = \{A, B\}$ and $\underline{S}_S = \emptyset$, the projection function $f^{U=1}{}_D$ for variable D is identical to the original function

f_D, so that an intervention on variable B setting $B = 0$ does not alter the value of variable D. With definition (EP AC), $B = 1$ is thus not an extended actual cause of $D = 1$.

Demonstrate now that $B = 1$ is an extended contributory cause of $D = 1$: Observe that for the construction of non-natural causal beams $S_D = \{A\}$with $\underline{S}_D = \{B\}$, and $S_D = \{B\}$ with $\underline{S}_D = \{A\}$ are admissible, namely by setting $W_D = \underline{S}_D = \{B\}$ with $B = 0$ in the first case and $W_D = \underline{S}_D = \{A\}$ and $A = 0$ in the second. In the resulting beam of the latter constellation, projection function $f^{U=1}{}_D$ has the form $f^{U=1}{}_D = id_B$, so that an intervention on B setting $B = 0$ brings about $D = 0$. As a result, $B = 1$ is verified as a contributory cause of $D = 1$ according to definition (EP CC).

With this, my extended Pearl account analyzes this case of contributory causation successfully, too, by reaching the same verdict as an application of Hitchcock's account, and more importantly by agreeing with our intuition in this case.

Summarizing, with this I certainly have not proven that my extended version of Pearl's account does not fare worse than Hitchcock's account in dealing with the classic three problems of Lewis's counterfactual theory. However, in my view I have given strong evidence for the following claims:

First, that my account of actual causation should be able to analyze cases containing an intransitivity of actual causation at least as well as Hitchcock's account can. Second, that my account of actual causation should also be able to deal with cases of an asymmetric over-determination of an effect by more than one cause, called preemption, at least as well as Hitchcock's account can. Third, that my account of contributory causation should be able to deal with cases of a symmetric over-determination of an effect by at least two causes working in a parallel way at least as well as Hitchcock's account can.

SUMMARY

What has this thesis achieved? I have described Judea Pearl's structural account of modeling token causation as presented in his book *Causality – Models, Reasoning, and Inference* as a milestone in causal modeling using counterfactual information. I have shown that there is evidence that this account of Pearl's, written from the perspective of an engineer, solves the classic three problems of Lewis's metaphysical counterfactual account of causation: Cases of symmetric and of asymmetric over-determination and the stipulation of the transitivity of causation. Moreover, I have argued for an understanding of Pearl's account of modeling token causation that splits the classic problem of analyzing causal relationships on the token level, with which Lewis dealt, into two sub-problems: first the generation of an appropriate model *M* to formalize a given situation *S*, and second the definition of token causation relative to such a model *M*. I have maintained that relativizing the analysis of token causation to a given model *M* provides a way to handle the fact that there usually is a multitude of equally justified descriptions of any one situation *S*, depending on which events and which kinds of events in *S* are of interest for us.

The focus in this thesis has clearly been on the second sub-problem of causal modeling. My main observations concerning Pearl's account and the second sub-problem are the following: Pearl's basic idea is that for an event *c* to qualify as a token cause for another event *e* in an arbitrary situation *S*, described by a model *M*, event *c* has to be able to sustain event *e* against certain contingencies. I have maintained that this can be most easily understood with the aid of the leading idea that *c* sustains its effect *e* via a causal process that links cause *c* and effect *e*. More precisely, for Pearl

if $M := \langle U,V,F \rangle$ in state $U=u$ is a causal model formalizing the actual situation S with $X=x$ formalizing event c and $Y=y$ formalizing event e, then x can be called an actual cause, or contributory cause respectively, of y according to M in the state u if and only if x sustains y against certain structural contingencies in a natural beam M_u, or in a non-natural causal beam M_u respectively. With this, I have reached the insight that a causal process can be understood to be the concatenation of the projection functions f_i^u in a causal or natural beam M_u of exactly the variables V_i that are intermediate between the cause variable X and the effect variable Y.

Moreover, I have drawn attention to the fact that values of exogenous variables U can neither be actual nor contributory causes for Pearl, mainly since he excludes these variables U from sustaining sets in the construction of both natural and causal beams M_u. For this reason, I have offered a natural extension of Pearl's account of actual and of contributory causation that allows interventions on exogenous variables. I have achieved this by slight modifications in Pearl's definitions of a causal and a natural beam, and of actual and contributory causation, by which a uniform treatment of exogenous and endogenous variables in constructing sustaining sets is facilitated. Finally, I have argued that Pearl's distinction between two mutually disjoint kinds of token causation – actual and contributory causation – is unnecessary, and I have offered a trivial unified definition of token causation that contains actual causation as a special case, although I have regarded this issue to be mainly a terminological problem

Concerning the first sub-problem of modeling token causation, with which I have only marginally dealt in this thesis, and Pearl's account, I have mainly made the following observations: Pearl does not directly deal with the problem of generating models M for a given situation S. His main concern is to generate causal models geared to analyzing causal relationships on the type level. Pearl's algorithm of inductive causation

starts out with a joint probability distribution P extending over a set $U \cup V$ and a definition of conditional probabilistic independence and generates an equivalence class of basic causal structures D. In this way, the result of Pearl's model generating procedure is not a complete causal model M with a set of deterministic functions F, but instead, the result is the equivalence class of basic causal structures D that are in essence just the set of variables $U \cup V$ together with a relation describing parenthood between these variables.

I have drawn the obvious conclusion that in the application of Pearl's definitions of actual or contributory causation, the causal model M out of which the beams M_u are constructed is strongly underdetermined, because the set of functions F determining the values of the variables in V has to be stipulated. I have interpreted this result as being compatible with the differentiation of causal analysis into two sub-problems: In the first step, probability distributions allow the extraction of causal patterns out of them ending up with a multitude of basic causal structures. In the second step, these basic causal structures can serve as blueprints for a complete causal model, relative to which token causation is defined.

Throughout this thesis, I have argued for an understanding of Hitchcock's and Woodward's work in 'The Intransitivity of Causation Revealed in Equations and Graphs' and in *Making Things Happen* as being an attempt by philosophers to give a formally less complicated modeling account of token causation that captures the leading idea of Pearl's account: A cause c and an effect e in a situation S are linked by a causal process that is a concatenation of mechanisms at work in S. I have given an elaborate corrective reconstruction of Hitchcock's account and have briefly touched on Woodward's account, which essentially takes over Hitchcock's definitions with a slightly changed terminology.

I have employed these accounts of Hitchcock and Woodward as a graphic introduction to Pearl's account, since the same notions of a causal model and of a causal process are defined in a formally more accessible way. In particular, I have motivated that Hitchcock and Woodward formalize a causal process between events c and e in a situation S by an active route, or a weakly active route respectively, in a causal model $\boldsymbol{M}$ between the variables X and Z whose actual values x and z formalize these events c and e. The only point in which my corrective reconstruction of Hitchcock's and Woodward's accounts has deviated from their originals has been in relativizing my reconstructed definitions of token causation in a situation S to a given model $\boldsymbol{M}$.

Concerning the problem of generating a causal model $\boldsymbol{M}$ to appropriately formalize a given situation S, I have tried to organize Hitchcock's fragmentary remarks in a coherent sketch that has consisted of two steps: First the extraction of variables in situation S, and second the establishment of counterfactual dependencies between these variables and construction of the structural equations $\boldsymbol{E}$. From this sketch I have drawn the conclusion that Hitchcock's structural equations are in essence another notation for exactly the counterfactual information that we put into the analysis of a situation S. Moreover, I have briefly motivated that Hitchcock's attempt to conceptually reduce token causation to ENF-counterfactuals fails, since the causal notion of a surgical intervention has to enter in the definition of an active route.

Finally, concerning the core question whether Hitchcock and Woodward succeed in giving a simplification of Pearl's modeling account of singular event causation, I have reached the result that they do not. I have motivated this result by constructing four paradigmatic examples to which I have applied my modeling reconstruction of Hitchcock's definitions of actual and contributory causation, my extension of Pearl's corresponding definitions that allow exogenous variables to be causes, and finally my modeling reconstruction of Woodward's definition of token causation.

Two of these examples have been geared to compare my modeling reconstruction of Hitchcock's definition of actual causation with my extension of Pearl's definition of actual causation allowing exogenous variables to be causes. For both examples, the verdicts that an analysis with my extension of Pearl's account has delivered have been in accord with what I take to be our causal intuitions, whereas the application of my reconstruction of Hitchcock's account has delivered results contradicting these intuitions in both cases.

The other two examples have been canonical extensions of the former, designed to compare my reconstruction of Hitchcock's definition of contributory causation with my extension of Pearl's definition of contributory causation. Again, for both examples, the verdicts that an analysis with my extension of Pearl's account has delivered have been in accord with what I take to be our causal intuitions. And the results that an application of my reconstruction of Hitchcock's account has delivered have still contradicted these intuitions in one of these cases.

An application of my modeling reconstruction of Woodward's inclusive account of token causation to these four examples has shown that Woodward's inclusive understanding of token causation has an advantage over Hitchcock's separated notions of actual and contributory causation. For, my modeling reconstruction of Woodward's account has analyzed two of my examples in a way that has been in accord with our causal intuition. Nevertheless, it has still failed in the analysis of the other two examples, where it has delivered results that have contradicted our intuition.

From the discussion of these four examples I have concluded that Hitchcock's and Woodward's project of simplifying Pearl's account fails because of their underlying rationale that identifies a causal process linking a cause and its effect with a route in a causal model. I have argued that this formalization of a process is too simplistic, lastly because it does not incorporate Pearl's notion of sustenance.

Finally, I have briefly dealt with Hitchcock's revocation of his account of token causation that he makes in a footnote in his later article 'A Tale of Two Effects'. This article implicitly features two new notions of component and net causation, and I have offered formalizations of token level reductions of these notions. Moreover, I have argued that these token level reductions cannot replace the basic notion of actual causation, so that I have unified them to a definition of merged actual causation. However, even this artificially reconstructed definition has failed in the analysis of one of my examples.

BIBLIOGRAPHY

[ASP 05] C. Avin, I. Shpitser, and J. Pearl (2005): 'Identifiability of Path-Specific Effects'. Technical Report *R-321*. Department of Computer Science, UCLA.

[CHP 04] J. Collins, N. Hall, and L. Paul. (eds.) (2004): *Causation and Counterfactuals*. Cambridge: MIT Press.

[jC 00] John Collins (2000): 'Preemptive Prevention'. *Journal of Philosophy*, vol. 97, pp. 223-234.

[nC 89] Nancy Cartwright (1989): *Nature's Capacities and Their Measurement*. Oxford: Clarendon Press.

[nC 99] Nancy Cartwright (1999): *The Dappled World*. Oxford: Oxford University Press.

[nC 02] Nancy Cartwright (2002): 'Against Modularity, the Causal Markov Condition and Any Link Between the Two: Comments on Hausman and Woodward'. *British Journal for the Philosophy of Science*, vol. 53, pp. 411-453.

[nC 03] Nancy Cartwright (2003): 'Two Theorems on Invariance and Causality'. *Philosophy of Science*, vol. 70, pp. 203-224.

[pD 00] Phil Dowe (2000): *Physical Causation*. Cambridge: Cambridge University Press.

[eE 91] Ellery Eells (1991): *Probabilistic Causality*. Cambridge: Cambridge University Press.

[ES 83] E. Eells and E. Sober (1983): 'Probabilistic Causality and the Question of Transitivity'. *Philosophy of Science*, vol. 50, pp. 35-57.

[GP 97] D. Galles and J. Pearl (1997): 'Axioms of Causal Relevance'. *Artificial Intelligence*, vol. 97, pp.9-43.

[GP 98] D. Galles and J. Pearl (1998): 'An Axiomatic Characterization of Causal Counterfactuals'. *Foundations of Science*, vol. 3, pp. 151-183.

[nG 79] Nelson Goodman (1979): *Fact, Fiction and Forecast*. Cambridge: Harvard University Press.

[nH 00] Ned Hall (2002): 'Causation and the Price of Transitivity'. *Journal of Philosophy*, vol. 97, pp. 198-222.

[nH 04] Ned Hall (2004): 'Two Concepts of Causation'. in [CHP 04], pp. 225-276.

[jH 00] J.Y. Halpern (2000): 'Axiomatic Causal Reasoning'. *Journal of Artificial Intelligence Research*, vol. 12, pp. 317-37.

[HP 00] J.Y. Halpern, J. Pearl (2000): 'Causes and Explanations: A Structural Model Approach – Part I: Causes'. Technical Report *R-266-UAI*. Computer Science Department, UCLA.

[HP 00b] J.Y. Halpern, J. Pearl (2000): 'Causes and Explanations: A Structural Model Approach – Part II: Explanations'. Technical Report *R-266-IJCAI*. Computer Science Department, UCLA.

[dH 88] Daniel Hausman (1988): *Causal Asymmetries*. New York: Cambridge University Press.

[dH 92] Daniel Hausman (1992): 'Thresholds, Transitivity, Overdetermination and Events'. *Analysis*, vol. 52, pp. 159-163.

[HW 99] D. Hausman and J. Woodward (1999): 'Independence, Invariance, and the Causal Markov Condition'. *British Journal for the Philosophy of Science*, vol. 50, pp. 521-583.

[HW 03] C. Hitchcock, J. Woodward (2003): 'Explanatory Generalizations, Part 1: A Counterfactual Account'. *Noûs*, vol. 37, pp. 1-24.

[HW 03b] C. Hitchcock, J. Woodward (2003): 'Explanatory Generalizations, Part 2: Plumbing Explanatory Depth'. *Noûs*, vol. 37, pp. 181-199.

[cH 93] Christopher Hitchcock: 'A Generalized Probabilistic Theory of Causal Relevance'. *Synthese*, vol. 97, pp. 335-364.

[cH 95] Christopher Hitchcock: 'The Mishap at Reichenbach Fall: Singular vs. General Causation'. *Philosophical Studies*, vol. 78, pp. 257-291.

[cH 01] Christopher Hitchcock (2001): 'The Intransitivity of Causation Revealed in Equations and Graphs'. *Journal of Philosophy*, vol. 98, pp. 273-299.

[cH 01b] Christopher Hitchcock (2001): 'A Tale of Two Effects'. *Philosophical Review*, vol. 110, pp. 361-396.

[cH 03] Christopher Hitchcock (2003): 'Of Humean Bondage'. *British Journal for the Philosophy of Science*, vol. 54, pp. 1-25.

[iK 01] Igal Kvart (2001): 'Counterexamples to Lewis's 'Causation as Influence''. *Australasian Journal of Philosophy*, vol. 79, pp. 411-423.

[iK 01b] Igal Kvart (2001): 'A Counterfactual Theory of Cause'. *Synthese*, vol.127, pp. 389-427.

[jK 73] Jaegwon Kim (1973): 'Causes and Counterfactuals'.
Journal of philosophy, vol. 70, pp. 570-572.

[jK 73b] Jaegwon Kim (1973): 'Causation, Nomic Subsumption, and the Concept of Event'.
Journal of philosophy, vol. 70, pp. 217-236.

[dL 73] David Lewis (1973): 'Causation'.
Journal of Philosophy, vol. 70, pp. 556-567.

[dL 73b] David Lewis (1973): *Counterfactuals*.
Cambridge: Harvard University Press.

[dL 79] David Lewis (1979): 'Counterfactual Dependence and Time's Arrow'.
Noûs, vol. 13, pp. 455-476.

[dL 86] David Lewis (1986): 'Postscripts to 'Causation''.
in *Philosophical Papers*, vol. II., pp. 172-213.
Oxford: Oxford University Press.

[dL 86b] David Lewis (1986): 'Causal Explanation'.
in *Philosophical Papers*, vol. II., pp. 214-240.
Oxford: Oxford University Press.

[dL 00] David Lewis (2000): 'Causation as Influence'.
Journal of Philosophy, vol. 97, pp. 182-197.

[MKT 97] V. McKim and S. Turner (eds.) (1997):
Causality in Crisis? Statistical Methods and the Search for Causal Knowledge in the Social Sciences.
Notre Dame: University of Notre Dame Press.

[jM 74] J.L. Mackie (1974): *The Cement of the Universe.*
Oxford: Clarendon Press.

[mM 95] Michael McDermott (1995): 'Redundant Causation'.
British Journal of the Philosophy of Science,
vol. 46, pp. 523-544.

[mM 95b] Michael McDermott (1995): 'Lewis on Causal Dependence'. *Australasian Journal of Philosophy*, vol. 73, pp. 129-139.

[pM 89] Peter Menzies (1989): Probabilistic Causation and Causal Processes: A Critique of Lewis'. *Philosophy of Science*, vol. 56, pp. 642-663.

[PV 91] J. Pearl and T. Verma (1991): 'A Theory of Inferred Causation'. in *Principles of Knowledge Representation and Reasoning: Proceedings of the 2nd International Conference.* ed. by J.A. Allen, R. Fikes and E. Sandewall, pp. 441-452. San Mateo: Morgan Kaufman.

[jP 88] Judea Pearl (1988): *Probabilistic Reasoning in Intelligent Systems.* San Mateo: Morgan Kaufman.

[jP 93] Judea Pearl (1993): 'On the Statistical Interpretation of Structural Equations'. Technical Report *R-200.* Department of Computer Science, UCLA.

[jP 98] Judea Pearl (1998): 'On the Definition of Actual Cause'. Technical Report *R-259.* Department of Computer Science, UCLA.

[jP 98b] Judea Pearl (1998): 'Graphs, Causality and Structural Equation Models'. *Sociological Methods and Research*, vol. 27, pp. 226-284.

[jP 00] Judea Pearl (2000): *Causality – Models, Reasoning and Inference.* Cambridge: Cambridge University Press.

[jP 01] Judea Pearl (2001): Direct and Indirect Effects. Technical Report *R-273.* Computer Science Department, UCLA.

[jP 04] Judea Pearl (2004): 'Robustness of Causal Claims'. Technical Report *R-320.* Computer Science Department, UCLA.

[lP 00] L.A. Paul (2000): 'Aspect Causation'. *Journal of Philosophy*, vol. 97, pp. 235-256.

[wS 84] W. Salmon (1984): *Scientific Explanation and the Causal Structure of the World.* Princeton: Princeton University Press.

[wS 94] W. Salmon (1994): 'Causality without Counterfactuals'. *Philosophy of Science*, vol. 61, pp. 297-312.

[jS 00] Jonathan Schaffer (2000): 'Trumping Preemption'. *Journal of Philosophy*, vol. 97, pp. 165-181.

[jS 00b] Jonathan Schaffer (2000): 'Causation by Disconnection'. *Philosophy of Science*, vol. 67, pp. 285-300.

[eS 85] Elliott Sober (1985): 'Two Concepts of Cause'. in *PSA 1984*, vol. 2, pp. 405-424. ed. by P. Asquith and P. Kitcher. East Lansing: Philosophy of Science Association.

[ST 93] E. Sosa and M. Tooley (eds.): *Causation.* Oxford: Oxford University Press.

[SGS 91] P. Spirtes, C. Glymour, and R. Scheines (1991): 'From Probability to Causality'. *Philosophical Studies*, vol. 64, pp. 1-36.

[SGS 00] P. Spirtes, C. Glymour, and R. Scheines (2000): *Causation, Prediction and Search*, 2nd ed.. New York: Springer Verlag.

[SRMSG 98] P. Spirtes, T. Richardson, C. Meek, R. Scheines, and C. Glymour (1998): 'Using Path Diagrams as a Structural Equation Modeling Tool.' *Sociological Methods and Research*, vol. 27, pp. 148-181.

[SSGMR 98] P. Spirtes, R. Scheines, C. Glymour, C. Meek and T. Richardson (1998): 'The TETRAD Project: Constraint Based Aids to Causal Model Specification'. *Multivariate Behavioral Research*, vol. 33, pp. 65-118.

[TP 02] J. Tian and J. Pearl (2002): 'A General Identification Condition for Causal Effects'. Technical Report *R-290*. Department of Computer Science, UCLA.

[jW 04] Jon Williamson (2004): *Bayesian Nets and Causality: Philosophical and Computational Foundations*. Oxford: Clarendon Press.

[jW 97] James Woodward (1997): 'Explanation, Invariance, and Intervention'. *PSA 1996*, vol. 2, pp. 26-41.

[jW 97b] James Woodward (1997): 'Causal Modeling, Probabilities and Invariance'. in [MKT 97], pp. 265-317.

[jW 99] James Woodward (1999): 'Causal Interpretation in Systems of Equations'. *Synthese*, vol. 121, pp. 199-257.

[jW 01] James Woodward (2001): 'Probabilistic Causality, Direct Causes, and Counterfactual Dependence'. in *Stochastic Causality*, pp. 39-63, ed. by M. Galavotti, P. Suppes, and D. Constantini. Stanford: CSLI.

[jW 02] James Woodward (2002): 'What is a Mechanism? A Counterfactual Account'. *Philosophy of Science*, vol. 69, pp. 366-377.

[jW 03] James Woodward (2003): *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.